THE UNIVERSITY
of EDINBURGH

Production of Plant Triterpenoids in the Yeast *Saccharomyces cerevisiae*

Matthew Peter Dale

Thesis presented for the degree of Doctor of Philosophy

Institute of Quantitative Biology, Biochemistry and Biotechnology

The University of Edinburgh

April 2019

Doctor of Philosophy – The University of Edinburgh – 2019

Declaration

I declare that this thesis was composed by myself and that the work presented is my own, except where otherwise stated. This thesis has not been submitted for any other degree or professional qualification.

Matthew Peter Dale

15th April 2019

Acknowledgements

I'd like to thank Susan Rosser for giving me the opportunity to undertake my PhD in her lab, and for her support throughout the process. I'd also like to thank Emily Johnston and Tessa Moses for their guidance and support during my project, and for reading and providing feedback on my thesis.

Thanks to Rebecca Holmes and Annegret Honsbein for helping me learn my way around working with yeast, and to everyone else in the lab for all your help throughout my project.

I'd like to thank Thomas Louveau for providing early help and protocols for the extraction and analysis of metabolites featured in this work. I'd also like to thank Hannah Florance and Eliane Salvo-Chirnside for helping me to develop these protocols and for training and support for running the GC-MS and LC-MS instruments. And thanks to Lisa Imrie for performing the proteomics analyses.

Thanks to Scott Neilson and Hille Tekotte for providing access to the Tecan Sunrise instruments used to perform the growth assays, and for training regarding the use of the BioLector.

Thanks also to my viva examiners Anne Osbourn and Stephen Wallace for valuable discussions and suggestions which have improved this thesis.

Finally, I'd like to my friends and family, and especially my father, for their support and encouragement throughout my PhD.

Abstract

The triterpenoids comprise a diverse family of plant natural products with potential applications in many sectors, including medicine, food, agriculture, and home and personal care. Triterpenoids are derived from the cyclisation of 2,3-oxidosqualene by an oxidosqualene cyclase (OSC), to generate a triterpene, which can then be oxidised by cytochromes P450 to produce triterpenoids, and glycosylated by UDP-glycosyltransferases (UGTs) to produce triterpenoid saponins. More than 150 triterpene structures have been identified to date, which can be modified in a myriad of ways, resulting in the huge diversity of triterpenoids found in nature. The modifications are often crucial for bioactivity. Numerous saponins have industrial potential as vaccine adjuvants, feeding deterrents, detergents and gelling agents, their amphipathic properties being critical to this activity. Meanwhile, many triterpenoid aglycones have potential as therapeutics (including anti-cancer, anti-HIV and hepatoprotective drugs) and insecticides. Many of these compounds are derived from oleanolic acid, which is produced from the triterpene β-amyrin by oxidation of carbon-28 (*C*-28; a methyl group) to a carboxyl group.

Despite this great potential, it is currently challenging to obtain triterpenoids in the quantities required for industrial exploitation. Plants typically accumulate triterpenoids in low abundance and under specific conditions, and many triterpenoids are produced by non-crop plants that are difficult to cultivate. Furthermore, plants often produce numerous structurally similar triterpenoids, making it difficult to purify the desired compound(s). Meanwhile, the complexity of triterpenoids, which contain multiple chiral centres and often undergo stereo- and regiospecific oxygenations, makes their chemical synthesis challenging and economically prohibitive. Triterpenoid production in the budding yeast *Saccharomyces cerevisiae* could be a means to address the shortfall in production. *S. cerevisiae* is a genetically tractable and well characterised microorganism that naturally produces 2,3-oxidosqualene, and is widely used in industrial fermentations for a variety of products. The present work focuses on the production of oleanane triterpenoids and saponins (i.e. derived from β-amyrin) in *S. cerevisiae*.

In Chapter 3, a gas chromatography-mass spectrometry (GC-MS) method to monitor and quantify the production of triterpenoids in yeast is presented.

In Chapter 4, twelve β-amyrin synthase (BAS) homologues are systematically compared for productivity in yeast, and a difference in β-amyrin production of > 10-fold was observed. The homologues from *Artemisia annua* (AaBAS) and *Chenopodium quinoa* (CqBAS1) were the most productive, each yielding 10.6 mg/L β-amyrin. Expression of most BAS homologues resulted in considerably slower growth indicative of metabolic burden. However, a BAS from *Avena strigosa* (AsBAS) had a negligible effect on growth while still producing a relatively high amount of β-amyrin (8.8 mg/L).

In Chapter 5, sixteen C-28 oxidase P450s (co-expressed with *AaBAS* and the cytochrome P450 reductase *ATR2*) are compared for the production of oleanolic acid. All strains grew slowly compared with a control strain carrying an empty vector. Product profiles varied considerably, and a 6.8-fold difference in oleanolic acid titre was observed. The CYP716AL1 enzyme from *Catharanthus roseus* produced the most oleanolic acid in the initial screen (14.1 mg/L), but also accumulated substantial amounts of β-amyrin (7.1 mg/L) and the intermediate compounds erythrodiol (6.7 mg/L) and oleanolic aldehyde (titre undetermined) compared with the other BASs. Co-expression of *CYP716AL1* with *AsBAS* resulted in faster growth and the production of mainly oleanolic acid, with very little β-amyrin, erythrodiol, or oleanolic aldehyde accumulating.

Finally, in Chapter 6, saponins derived from β-amyrin and oleanolic acid are produced in yeast through the additional expression of UGT enzymes. This study identified glycosylations at different positions on the triterpenoid backbone, and is the first reported production of an oleanane diglycoside saponin in yeast.

Lay summary

The triterpenoids are a group of plant chemicals with potential applications in medicine, agriculture, cosmetics and numerous other sectors. Triterpenoids are traditionally obtained by extraction from the plants that naturally produce them. However, plants usually produce very low amounts of triterpenoids, while high quantities are required for commercial exploitation. As such, an alternative method of production is required if triterpenoids are to be fully exploited for commercial and medical uses.

One such method is to utilise microorganisms such as baker's yeast. Triterpenoids are produced through a series of chemical reactions, and plants contain sets of genes that enable them to perform these reactions and produce triterpenoids. These genes can be introduced into yeast to produce triterpenoids. This approach has the potential to produce much higher quantities of triterpenoids than can be obtained from their natural sources. This is possible because, unlike plants, yeast cultures can be grown in large fermenters. However, this method is currently in its infancy and much work is required before commercially-relevant quantities can be produced in yeast. In addition, many triterpenoids have yet to be produced in yeast due to their structural complexity.

The present work attempts to address some of the current challenges in this field. In particular, two key reaction steps were optimised in order to boost triterpenoid production in yeast. Furthermore, a set of structurally complex triterpenoids whose production in yeast had not been reported before were produced.

Table of Contents

List of Figures

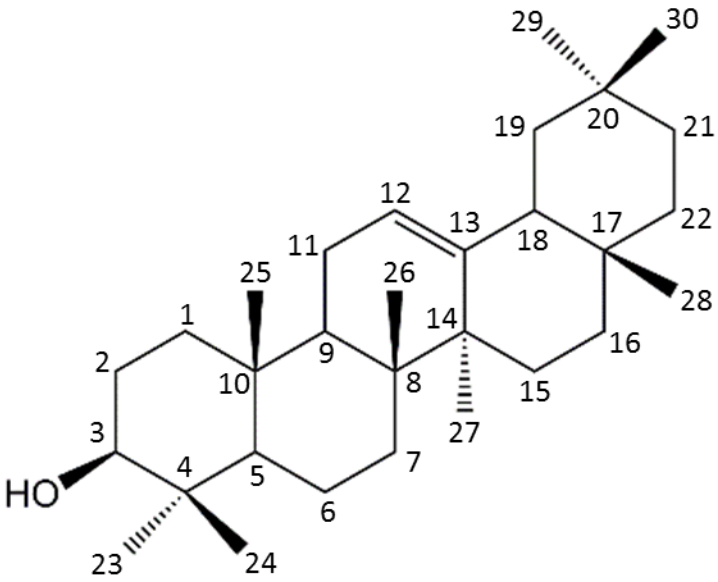
List of Tables

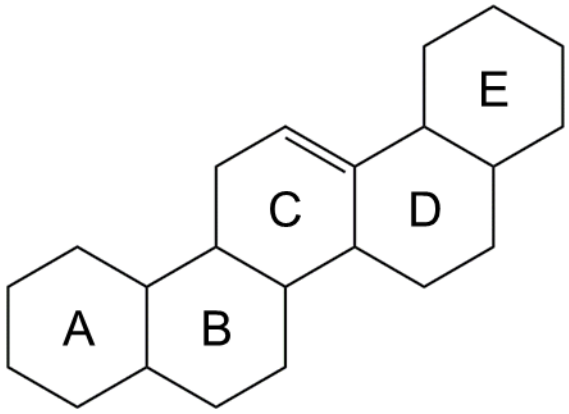
Abbreviations

BAS	β-amyrin synthase
CBC	Chair-boat-chair conformation
CCC	Chair-chair-chair conformation
CPR	NADPH-cytochrome P450 reductase
DMAPP	Dimethylallyl pyrophosphate
EI	Electron ionisation
EIC	Extracted ion chromatogram
ESI	Electrospray ionisation
FPP	Farnesyl pyrophosphate
Gal	Galactose
GC-MS	Gas Chromatography-Mass Spectrometry
Glc	Glucose
GPP	Geranyl pyrophosphate
HPLC	High-performance liquid chromatography
IPP	Isopentenyl pyrophosphate
IS	Internal standard
LC-MS	Liquid Chromatography-Mass Spectrometry
LUS	Lupeol synthase
MEV	Mevalonate pathway
OD	Optical density
OSC	Oxidosqualene cyclase
P450	Cytochrome P450
rDA	Retro-Diels Alder
RT-PCR	Reverse-transcription PCR
TIC	Total ion chromatogram
TLC	Thin-Layer Chromatography
UDP	Uridine diphosphate
UGT	UDP-glycosyltransferase

Triterpene numbering

The numbering convention for oleanane triterpenes (β-amyrin and its derivatives) is shown below. Carbons are numbered one to thirty, and the rings are labelled A to E. The triterpene shown is β-amyrin.





Chapter 1: Introduction

The triterpenoids are a large family of plant natural products with diverse bioactivities. In plants, they are important for resistance to pathogens and pests, with many having antimicrobial and anti-feedant activities (Moses et al., 2014a), and may also be involved in responses to abiotic stresses, such as nutrient starvation, light and temperature levels, and dehydration (Moses et al., 2014a; Thimmappa et al., 2014). Furthermore, triterpenoids have also been implicated in plant growth and development (Thimmappa et al., 2014). Due to their diverse bioactivities, triterpenoids have applications in numerous sectors, including in medicine, food, agriculture, and home and personal care (Moses et al., 2014a, 2013; Osbourn et al., 2011). They are derived from the cyclisation of 2,3-oxidosqualene, which is also a precursor to the sterols. Specifically, the cyclisation of 2,3-oxidosqualene by an oxidosqualene cyclase (OSC) generates a triterpene, and many different OSCs and triterpene cyclisation products have been identified to date (Hoshino, 2017; Ghosh, 2016; Thimmappa et al., 2014). The triterpenes can subsequently be modified by various enzymes to produce triterpenoids, resulting in the huge structural diversity that explains the diverse properties and applications of this family of metabolites (Moses et al., 2014a; Thimmappa et al., 2014).

Unfortunately, current methods of obtaining triterpenoids have severe limitations that often make them unsuitable for commercial exploitation, which requires triterpenoids to be produced in high quantities and purities at relatively low cost (Moses et al., 2013). Triterpenoids are traditionally obtained by extraction from plants, which can be low-yielding, time-consuming, and expensive (Moses et al., 2013). Furthermore, the structural complexity of triterpenoids makes their chemical synthesis challenging (Zhu et al., 2008; Cheng et al., 2009; Xiao and Yu, 2013; Deng et al., 1999; Kim et al., 2013). An alternative method is to produce triterpenoids in a microorganism expressing the appropriate biosynthetic pathway, and the yeast *Saccharomyces cerevisiae* has been successfully used for this purpose (Kirby et al., 2008; Fukushima et al., 2013; Moses et al., 2014b; Czarnotta et al., 2017; Zhuang et al., 2017a; Arendt et al., 2017; Zhao et al., 2018; Zhu et al., 2018; Zhao et al., 2017). In this chapter, the biosynthesis, properties, and methods of obtaining triterpenoids, with a focus on their biotechnological production in yeast, are discussed.

1.1 Triterpenoid biosynthesis in plants

All terpenoids are synthesised from the five-carbon molecules isopentenyl pyrophosphate (IPP) and dimethylallyl pyrophosphate (DMAPP) (Fig. 1.1) (Kirby and Keasling, 2009). The end-to-end joining of IPP and DMAPP, catalysed by prenyltransferases, produces linear terpene backbones that contain multiples of five carbons (Figure 1.1), and these can be cyclised by a variety of terpene synthases to produce cyclic terpenes (Kirby and Keasling, 2009). A terpene is thus defined by the number of carbons it contains, and those containing 30 carbons are called triterpenes.

Triterpenoids are modified derivatives of triterpenes, polycyclic terpenes that are produced from the cyclisation of the 30-carbon 2,3-oxidosqualene by an oxidosqualene cyclase (OSC) (Figure 1.2, Figure 1.3) (Thimmappa et al., 2014). Different triterpenes can be produced depending on the OSC (Figure 1.3), with more than 100 OSCs and 150 different triterpenes having been identified to date (Hoshino, 2017; Ghosh, 2016). Following cyclisation, the resulting triterpenes can be modified by a variety of enzymes. The most common modifications are oxygenations catalysed by cytochromes P450 (producing triterpenoids; Figure 1.4), and glycosylations catalysed by UDP-glycosyltransferases (UGTs) (producing triterpenoid saponins; Figure 1.5) (Moses et al., 2014a; Thimmappa et al., 2014). These modifications can occur at various positions along the numerous triterpene backbones, resulting in the vast diversity of triterpenoid structures and bioactivities found in nature (Moses et al., 2014a).

This section provides an overview of triterpenoid biosynthesis, from the generation of 2,3-oxidosqualene and its cyclisation to a triterpene, to the decoration of triterpene backbones by P450s and UGTs to produce triterpenoids and saponins.

1.1.1 Generation of 2,3-oxidosqualene

In plants, IPP and DMAPP are synthesised via two pathways: (i) the mevalonate (MEV) pathway, which is localised to the cytosol, and (ii) the 1-deoxy-D-xylulose-5-phosphate (DXP) pathway, which is localised to the plastids (Kirby and Keasling, 2009). In general, eukaryotes and archaea possess the MEV pathway (with plants also having the DXP pathway), while bacteria typically possess the DXP pathway (Kirby and Keasling, 2009).

Plants produce triterpenoids via the MEV pathway (Kirby and Keasling, 2009). Here, acetyl-CoA is converted to IPP through a series of six reactions, in which mevalonate is an intermediate (Figure 1.1) (Miziorko, 2011). First, the condensation of two acetyl-CoA molecules by acetyl-CoA thiolase generates acetoacetyl-CoA. Subsequently, the addition of another acetyl-CoA molecule by HMG-CoA synthase generates 3-hydroxy-3-methylglutaryl-CoA (HMG-CoA), which is then reduced to mevalonate in an NADPH-dependent reaction catalysed by HMG-CoA reductase (HMGR). The phosphorylation of mevalonate by mevalonate kinase then generates mevalonate-5-phosphate, which is further phosphorylated by phosphomevalonate kinase to produce mevalonate-5-pyrophosphate. Finally, mevalonate-5-pyrophosphate is converted to IPP in a decarboxylation reaction catalysed by mevalonate pyrophosphate decarboxylase (Miziorko, 2011).

After the generation of IPP, the enzyme IPP-DMAPP isomerase (IDI) catalyses the reversible isomerisation between IPP and DMAPP (Figure 1.1) (Moses et al., 2014a; Kirby and Keasling, 2009). These two molecules are combined by the prenyltransferase geranyl pyrophosphate synthase to produce the 10-carbon geranyl pyrophosphate (GPP) (Kirby and Keasling, 2009). Subsequently, the addition of another IPP molecule to GPP by farnesyl pyrophosphate synthase produces the 15-carbon farnesyl pyrophosphate (FPP), and the fusion of two FPP molecules by squalene synthase produces the 30-carbon squalene (Moses et al., 2014a; Kirby and Keasling, 2009). This is then oxidised to 2,3-oxidosqualene by squalene epoxidase, which introduces an epoxide group (Figure 1.1) (Moses et al., 2014a), and the cyclisation of 2,3-oxidosqualene produces a triterpene (30-carbon terpene) (Thimmappa et al., 2014). Similarly, the cyclisation of GPP and FPP produces monoterpenes and sesquiterpenes (10- and 15-carbon terpenes), respectively (Kirby and Keasling, 2009). In the biosynthesis of diterpenes (20 carbons), the addition of IPP to FPP generates the 20-carbon geranylgeranyl pyrophosphate (GGPP), which is then cyclised to a diterpene (Kirby and Keasling, 2009). Meanwhile, the fusion of two GGPP molecules produces the 40-carbon tetraterpenes (Nisar et al., 2015).

1.1.2 Cyclisation of 2,3-oxidosqualene

The OSC-catalysed cyclisation of 2,3-oxidosqualene to a triterpene is the first committed step in triterpenoid biosynthesis (Figure 1.2) (Thimmappa et al., 2014). Sterols are also derived from the cyclisation 2,3-oxidosqualene by an OSC

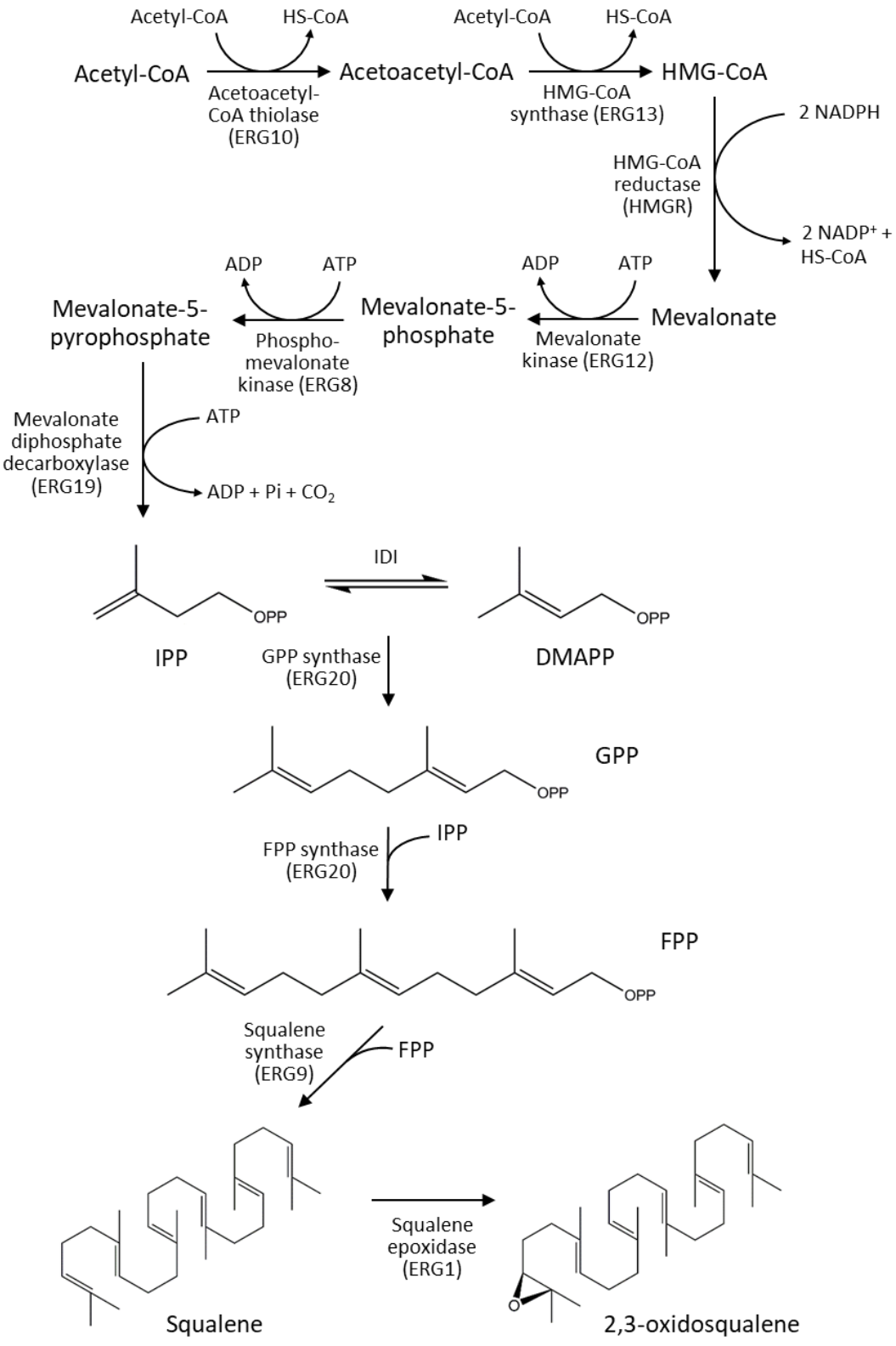


Figure 1.1. Generation of 2,3-oxidosqualene from acetyl-CoA via the mevalonate pathway. See next page for full legend.

(Figure 1.2) (Thimmappa et al., 2014); in animals and fungi, lanosterol synthase cyclises 2,3-oxidosqualene to lanosterol, which is then converted to the essential membrane sterols cholesterol (in animals) or ergosterol (in fungi) (Thimmappa et al., 2014). In plants, cycloartenol synthase converts 2,3-oxidosqualene to cycloartenol, a precursor to the phytosterols, although lanosterol also occurs in plants (Thimmappa et al., 2014; Moses et al., 2014a). In general, sterol and triterpene biosynthesis can be distinguished by the conformations of the first three rings during cyclisation. In sterol biosynthesis they take on a chair-boat-chair (CBC) conformation, while in triterpene biosynthesis they have a chair-chair-chair (CCC) conformation (Figure 1.2) (Thimmappa et al., 2014). The cyclisation of 2,3-oxidosqualene thus represents an important branch point between triterpenoid and sterol biosynthesis.

During cyclisation the OSC first binds 2,3-oxidosqualene, which folds into either the CBC or CCC conformation, depending on whether the OSC is a sterol or triterpene synthase, respectively (Thimmappa et al., 2014). The catalytic aspartate residue of the OSC then protonates the epoxide group of 2,3-oxidosqualene, which initiates cyclisation (Figure 1.3) (Hoshino, 2017). Cyclisation involves numerous ring generations, ring expansions and skeletal rearrangements, and proceeds through a series of carbocation intermediates (Hoshino, 2017; Thimmappa et al., 2014). Deprotonation or water capture of an intermediate terminates the reaction and generates a triterpene, and the intermediate at which the reaction was terminated determines which triterpene is produced (Figure 1.3). This explains the huge diversity of triterpenes that occur in nature, as there are many possible carbocation intermediates generated during the cyclisation reaction and each leads to the production of a different triterpene (Thimmappa et al., 2014).

Figure 1.1. Generation of 2,3-oxidosqualene from acetyl-CoA via the mevalonate pathway. The mevalonate pathway converts acetyl-CoA to the five-carbon molecule isopentenyl pyrophosphate (IPP) through six reactions, where mevalonate is an intermediate. Subsequently, IPP is isomerised to dimethylallyl pyrophosphate (DMAPP). The end-to-end joining of IPP and DMAPP generates geranyl pyrophosphate (GPP; 10 carbons) then farnesyl pyrophosphate (FPP; 15 carbons), and the joining of two FPP molecules generates the 30-carbon squalene. Finally, epoxidation of squalene produces 2,3-oxidosqualene, the precursor to the triterpenes. Names of the corresponding *S. cerevisiae* genes are shown in parentheses.

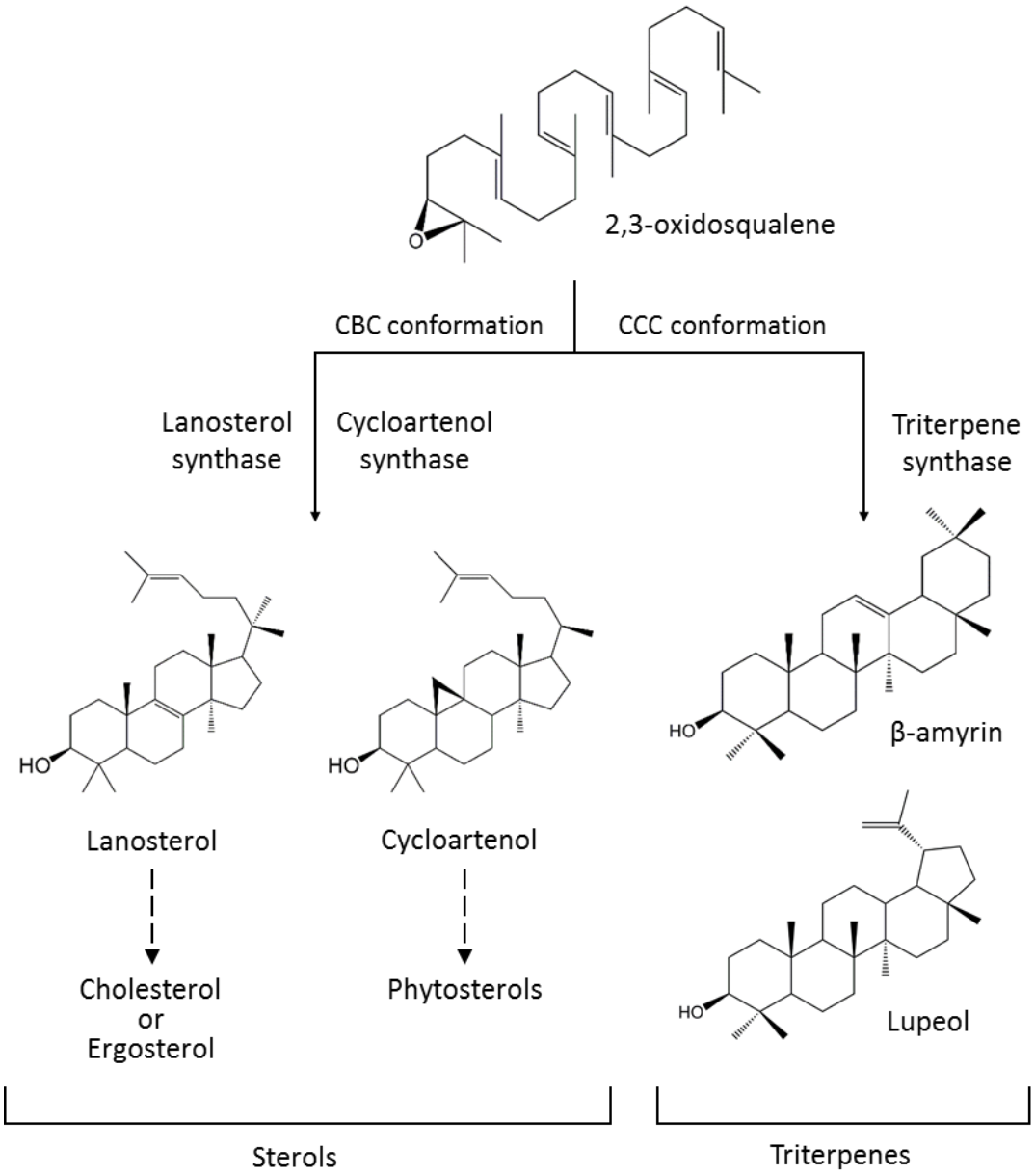


Figure 1.2. Cyclisation of 2,3-oxidosqualene to sterols and triterpenes. In the formation of sterols, 2,3-oxidosqualene folds into the chair-boat-chair (CBC) conformation and is cyclised by a lanosterol synthase or a cycloartenol synthase (left). Lanosterol is the precursor to the essential membrane lipids cholesterol (in animals) and ergosterol (in fungi). Cyclisation to a triterpene proceeds *via* the chair-chair-chair (CCC) conformation (right). Many OSCs catalysing the cyclisation of 2,3-oxidosqualene to a variety of triterpenes have been identified; two such triterpenes (β-amyrin and lupeol) are shown here. Dashed arrows represent multiple reactions.

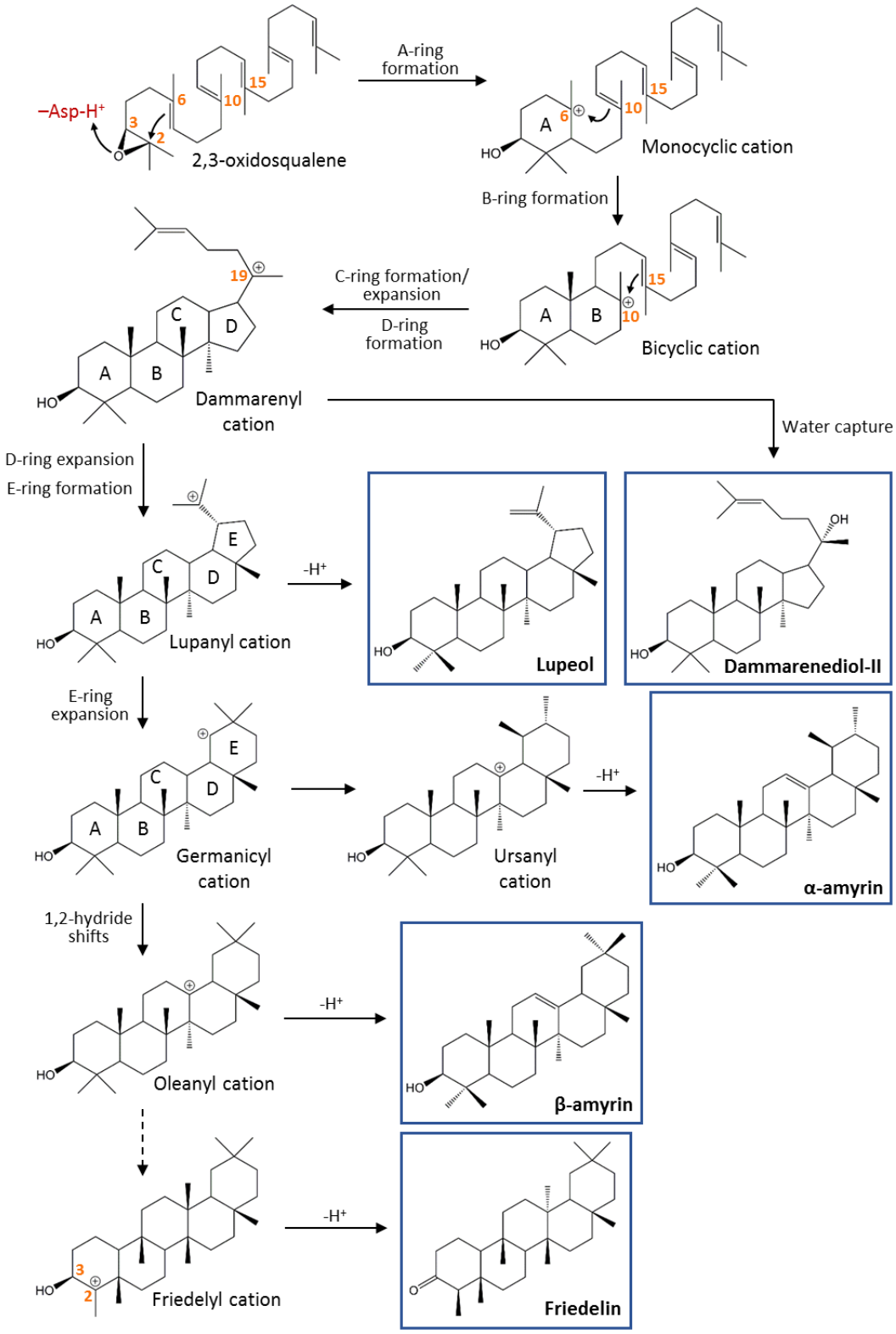


Figure 1.3. Oxidosqualene cyclisation cascade to generate triterpenes. See next page for legend.

Protonation of 2,3-oxidosqualene results in the generation of the six-membered A-ring with a positive charge at carbon-6 (*C*-6; 2,3-oxidosqualene numbering) (Figure 1.3). This generates a monocyclic cation, which is the first carbocation intermediate in the cyclisation cascade (Hoshino, 2017). The A-ring contains the hydroxyl group that is derived from the epoxide of 2,3-oxidosqualene. Subsequently, the six-membered B-ring forms, generating a bicyclic cation with a positive charge at *C*-10 (Hoshino, 2017). Rings C to E are then formed sequentially, and each is generated through the initial formation of a five-membered ring followed by ring expansion to a six-membered ring (Hoshino, 2017). The ring forming reactions result in the further movement of the positive charge across the molecule. This, combined with additional rearrangements in the form of 1,2-hydride and 1,2-methyl shifts, generates a variety of carbocation intermediates having the positive charge and methyl groups at different positions (Hoshino, 2017; Thimmappa et al., 2014). Thus, a range of carbocation intermediates, with different numbers and sizes of rings, different positions of methyl groups, and with the positive charge at different positions, are generated throughout the cyclisation cascade (Figure 1.3) (Hoshino, 2017; Thimmappa et al., 2014).

Termination of the reaction through deprotonation or water capture results in the formation of a double bond or a second hydroxyl group, respectively, at the site of the positive charge, and generates a triterpene (Figure 1.3) (Thimmappa et al., 2014; Salmon et al., 2016). For example, the dammarenyl cation is generated after the formation of the five-membered D-ring, and is therefore a 6/6/6/5 tetracycle (i.e. containing three six-membered rings and one five-membered ring). Water capture at *C*-19 (2,3-oxidosqualene numbering), which holds the positive charge, results in the

Figure 1.3. Oxidosqualene cyclisation cascade to generate triterpenes. After binding the OSC, 2,3-oxidosqualene folds into the chair-chair-chair (CCC) conformation. Protonation of the epoxide group by the catalytic aspartate (red) initiates cyclisation, which proceeds through a series of carbocation intermediates formed through ring generations, ring expansions, and 1,2-hydride and -methyl shifts. Deprotonation of a carbocation intermediate by the OSC (or occasionally water capture) terminates the cyclisation cascade and generates a triterpene. Different triterpenes can be produced depending at which intermediate the reaction is terminated. Formation of the triterpenes dammarenediol-II, lupeol, α-amyrin and β-amyrin, as well as the highly rearranged triterpene friedelin, are shown (blue boxes). Carbon atoms are labelled according to 2,3-oxidosqualene numbering (orange), and rings A through E are labelled. The dashed arrow represents multiple reaction steps.

addition of a second hydroxyl group and termination of the reaction, generating the tetracyclic triterpene dammarenediol-II (Figure 1.3) (Hoshino, 2017; Salmon et al., 2016). Alternatively, ring expansion of the dammarenyl cation D-ring to a six-membered ring, followed by generation of the five-membered E-ring, gives the 6/6/6/6/5 pentacyclic lupanyl cation, deprotonation of which generates the triterpene lupeol (Figure 1.3) (Thimmappa et al., 2014; Hoshino, 2017). The widespread triterpene β-amyrin is produced from deprotonation of the oleanyl cation, which is generated through ring expansion of the lupanyl cation E-ring (giving the 6/6/6/6/6 pentacyclic germanicyl cation) followed by two 1,2-hydride shifts (Figure 1.3) (Thimmappa et al., 2014; Hoshino, 2017). Alpha-amyrin, which differs from β-amyrin only in the position of a methyl group, is derived from the ursanyl cation, whose formation from the germanicyl cation involves a 1,2-methyl shift (Figure 1.3) (Thimmappa et al., 2014). The most highly rearranged carbocation intermediate is the friedelyl cation, whose formation involves ten 1,2 shifts from the dammarenyl cation. Reaction termination at this intermediate produces the triterpene friedelin and occurs by nucleophilic attack of the positive charge at *C*-2 by the *C*-3 hydroxyl group. This generates a ketone at *C*-3, and friedelin thus has no C-C double bond (Figure 1.3) (Thimmappa et al., 2014).

Oxidosqualene cyclases can be classed as monofunctional or multifunctional. Monofunctional OSCs produce a single triterpene (usually with trace amounts of other triterpenes), while multifunctional OSCs generate multiple triterpene cyclisation products, often β-amyrin, α-amyrin and/or lupeol (Thimmappa et al., 2014; Ghosh, 2016). For example, the *Arabidopsis thaliana* baruol synthase produces 23 different triterpenes, with baruol as its major product (Shan et al., 2005). By contrast, examples of monofunctional OSCs include numerous β-amyrin and lupeol synthases (Thimmappa et al., 2014). The existence of multifunctional OSCs can be understood by considering the nature of the cyclisation reaction. Failure to terminate at a particular carbocation intermediate would allow cyclisation to continue, resulting in termination at a different intermediate and the production of a different triterpene (Figure 1.3) (Thimmappa et al., 2014). Thus, OSCs that are highly effective at terminating the reaction at a specific carbocation intermediate are monofunctional, while those that are less effective at this produce multiple triterpenes and so are multifunctional.

1.1.3 Decoration of the triterpene scaffold to produce triterpenoids and triterpenoid saponins

Following cyclisation, the triterpene can be modified by various enzymes in decoration (or tailoring) reactions. The most common modifications are oxygenations catalysed by P450s (Figure 1.4) and glycosylations catalysed by UGTs (Figure 1.5) (Moses et al., 2014a; Thimmappa et al., 2014). Other modifications include acylations, methylations and malonylations (Moses et al., 2014a). Such modified triterpenes containing heteroatoms are called triterpenoids, and in the case of a glycosylated triterpene/triterpenoid the term triterpenoid saponin is used. These modifications have a substantial impact on the properties of the triterpenoid and are crucial for bioactivity (for more information see section 1.2). Oxygenations, glycosylations and other types of tailoring reactions are discussed in this section.

1.1.3.1 Cytochrome P450-catalysed oxygenations to produce triterpenoids

Cytochromes P450 are a superfamily of monooxygenases that catalyse diverse reactions, often stereo- and regiospecific oxygenations (e.g. hydroxylations) (Bak et al., 2011). They are found in all kingdoms of life (Sezutsu et al., 2013) and are especially abundant in plants, where they represent the largest family of biosynthetic enzymes (Nelson and Werck-Reichhart, 2011). They are involved in a host of plant processes, from the detoxification of xenobiotics, to the biosynthesis of signalling hormones and the production of secondary metabolites (Bak et al., 2011), and are extensively involved in triterpenoid biosynthesis (Thimmappa et al., 2014). Catalysis involves the reduction of molecular oxygen at the active site haem group to generate a reactive oxygen radical, which then oxygenates the substrate (Werck-Reichhart and Feyereisen, 2000). Plant P450s receive the electrons for this reaction through interaction with an NADPH-cytochrome P450 reductase (CPR), which accepts electrons from NADPH and shuttles them through its FMN and FAD domains to the P450 haem group (Werck-Reichhart and Feyereisen, 2000; Bak et al., 2011; Jensen and Møller, 2010). Plant P450s thus require molecular oxygen, the co-factors NADPH, FMN, FAD and haem, and the co-expression of a CPR for activity.

Numerous P450s involved in triterpenoid biosynthesis have been identified (Ghosh, 2017; Miettinen et al., 2017). They are highly regio- and stereospecific, usually catalysing oxygenations that introduce a hydroxyl, ketone/aldehyde, or carboxyl group to various positions on the triterpene backbone (Augustin et al., 2011; Thimmappa et

al., 2014). Epoxidation has also been reported (Geisler et al., 2013; Miettinen et al., 2017), as have non-oxygenation reactions such as desaturations and carbon-carbon bond cleavage (Field and Osbourn, 2008; Castillo et al., 2013). Many functional groups introduced by P450s can be targeted for further modification, such as by UGTs or acyltransferases (Thimmappa et al., 2014; Moses et al., 2014a). In addition, multiple P450s can act on the same triterpene backbone to generate more highly oxygenated triterpenoids (Han et al., 2018; Seki et al., 2011; Moses et al., 2014b; Arendt et al., 2017). While most reported P450s involved in triterpenoid biosynthesis are monofunctional, catalysing a single reaction, some catalyse multiple reactions. For example, CYP51H10 from *Avena strigosa* introduces both a hydroxyl and an epoxide group to β-amyrin (Geisler et al., 2013), while CYP716A2 from *A. thaliana* hydroxylates β-amyrin at the *C*-16, *C*-22 or *C*-28 positions to generate three different products (Figure 1.4) (Yasumoto et al., 2016). To date, more than fifty P450s involved in triterpenoid biosynthesis have been identified, oxidising numerous positions across the triterpene backbone (Ghosh, 2017).

Cytochromes P450 are grouped into families and subfamilies based on amino acid sequence identity, with family members typically having ≥ 40 % amino acid sequence identity and subfamily members ≥ 55 % sequence identity (Bak et al., 2011). They are named systematically after their family and subfamily by a nomenclature committee (http://drnelson.uthsc.edu/CytochromeP450.html) (Nelson, 2009). For example, CYP51H10 is the tenth member of family 51, subfamily H. To date, P450s involved in triterpenoid biosynthesis have been identified across 10 families and 17 subfamilies (Ghosh, 2017). Cytochromes P450 are also divided into clans, where a clan is simply a clade of P450s (Werck-Reichhart and Feyereisen, 2000). Each clan thus contains one or multiple P450 families, and clans are named after the lowest-numbered P450 family that they contain (Werck-Reichhart and Feyereisen, 2000). For example, the CYP85 clan contains several families including CYP85, CYP88 and CYP716, and is named CYP85 as this is the lowest-numbered family. Land plants contain 11 clans (Nelson and Werck-Reichhart, 2011) and P450s involved in triterpenoid biosynthesis are found in five of these clans (Miettinen et al., 2017).

Many P450s involved in triterpenoid biosynthesis belong to the CYP716 family (Ghosh, 2017). In particular, members of the CYP716A subfamily typically catalyse the *C*-28 oxygenation of pentacyclic triterpenes such as the β-amyrin, α-amyrin and

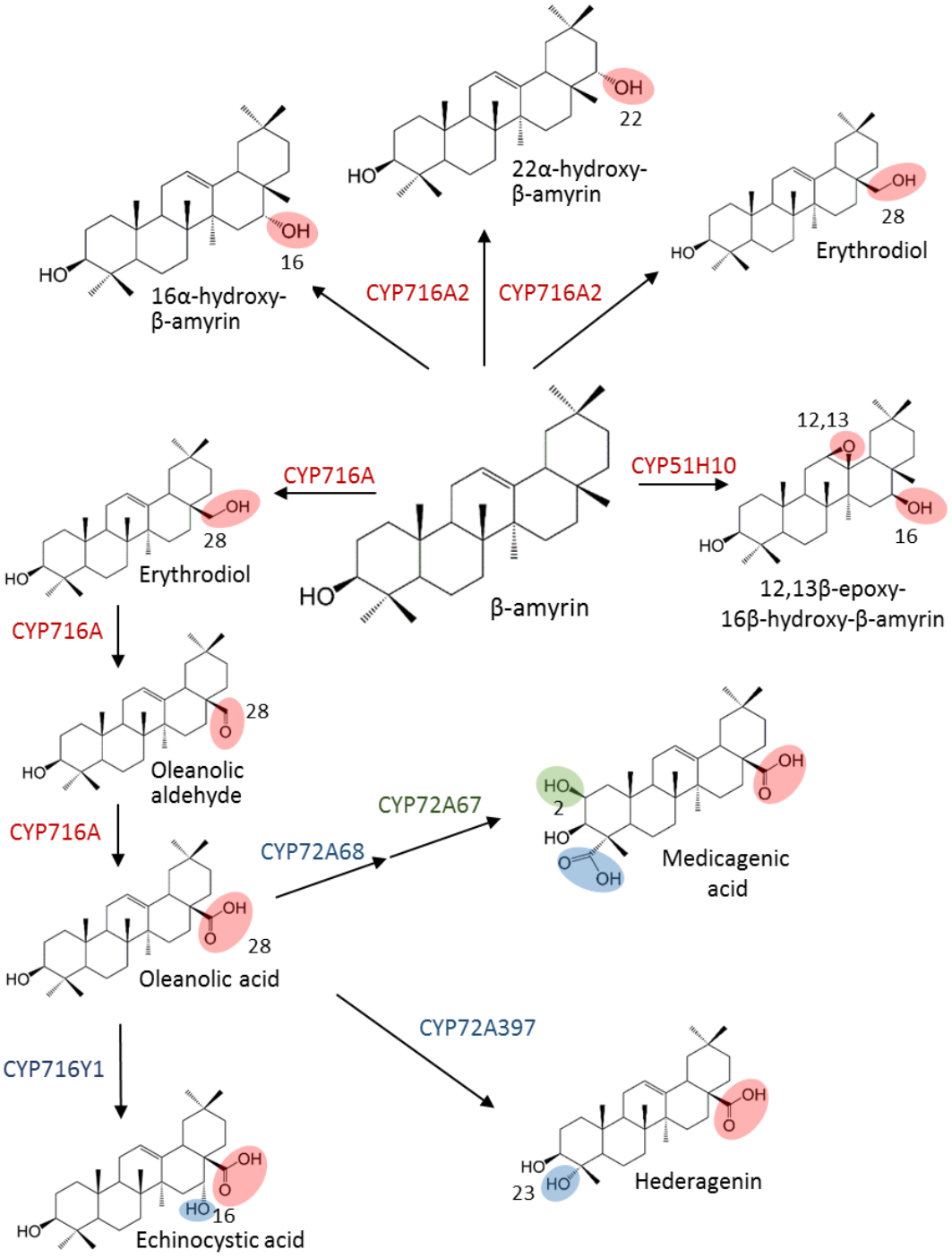


Figure 1.4. Oxygenation of β-amyrin by cytochromes P450 generates diverse triterpenoids. The P450-catalysed oxygenation of a triterpene such as β-amyrin can occur at multiple positions, generating a variety of triterpenoid products (red). These triterpenoids can be further oxygenated by additional P450s (blue and green), resulting in a huge diversity of triterpenoid products.

lupeol, oxidising the *C*-28 methyl group to a carboxyl group (Figure 1.4) (Miettinen et al., 2017; Seki et al., 2011; Fiallos-Jurado et al., 2016; Khakimov et al., 2015; Yasumoto et al., 2017; Moses et al., 2015b, 2015a; Yasumoto et al., 2016). The *C*-28 carboxyl is a particularly important functional group because it is a common site of glycosylation (Thimmappa et al., 2014), and numerous bioactive triterpenoids and saponins contain this moiety (see section 1.2 for more information) (Nielsen et al., 2010; Ragupathi et al., 2011; Smith et al., 2007; Pollier and Goossens, 2012; Tsao et al., 2010; Suh et al., 2012; Tong et al., 2017). This oxygenation reaction appears to occur through intermediates containing a hydroxyl or aldehyde group at *C*-28 (Figure 1.4). For example, expression of the *C*-28 oxygenase CYP716A12 in a yeast strain engineered to produce β-amyrin resulted in the production of erythrodiol, oleanolic aldehyde and oleanolic acid, which are the *C*-28 hydroxy, aldehyde and carboxyl derivatives of β-amyrin, respectively (Figure 1.4) (Fukushima et al., 2011). While most CYP716As catalyse *C*-28 oxygenation, others such as CYP716A2 (a β-amyrin C-16, C-22 and C-28 hydroxylase) catalyse different reactions (Figure 1.4) (Miettinen et al., 2017; Ghosh, 2017; Yasumoto et al., 2016). The CYP716 family contains several other subfamilies that catalyse a range of oxygenations against tetracyclic and pentacyclic triterpenes (Figure 1.4), but only the CYP716As catalyse *C*-28 oxygenation (Miettinen et al., 2017).

1.1.3.2 UGT-catalysed glycosylations to produce triterpenoid saponins

Triterpenoids are commonly glycosylated to produce triterpenoid saponins (Augustin et al., 2011; Moses et al., 2014a). Glycosylation is performed by family 1 UDP-glycosyltransferases (UGTs), which use UDP-sugar donors to glycosylate their substrates (Thimmappa et al., 2014). Sugars are added as monosaccharides in a sequential fashion, rather than as previously synthesised oligosaccharides (Thimmappa et al., 2014; Shibuya et al., 2010; Wang et al., 2015), and the most common sugar donors in saponin biosynthesis include UDP-glucose, -galactose, -rhamnose, -arabinose, -xylose, and -glucuronic acid (Thimmappa et al., 2014; Augustin et al., 2011).

Glycosylations typically occur at the hydroxyl or carboxyl groups of the triterpenoid, most commonly at the *C*-3 hydroxyl and *C*-28 carboxyl groups for pentacyclic triterpenoids (such as those derived from β-amyrin) (Figure 1.5) (Thimmappa et al., 2014). While the *C*-3 hydroxyl is present in unmodified triterpenes, being derived from

the epoxide group of 2,3-oxidosqualene, the *C*-28 carboxyl must be introduced by a P450 (see above; Figure 1.4). Glycosylation can also less commonly occur at other oxygenated positions on the backbone of pentacyclic triterpenoids (Augustin et al., 2011; Sayama et al., 2012). By contrast, the ginsenosides, saponins that are derived from the tetracyclic dammarane-type triterpenoids, are commonly glycosylated at the *C*-3, *C*-6 and *C*-20 hydroxyl groups (Wang et al., 2015; Wei et al., 2015; Yan et al., 2014), of which the *C*-6 hydroxyl must be introduced by a P450 (Han et al., 2012). Thus, P450s, particularly those of the CYP716A subfamily that introduce the *C*-28 carboxyl group, are important for the biosynthesis of saponins, introducing the additional reactive oxygen moieties that can be glycosylated by UGTs (Moses et al., 2014a).

Saponins can contain one or multiple sugar chains, which can be linear or branched, and each sugar chain usually contains between two and five monosaccharides (Augustin et al., 2011). The sugar chains can be subject to further modifications, such as acylations (Ragupathi et al., 2011). Saponins containing a single sugar chain are monodesmosidic, while those with two sugar chains are bidesmosidic, and saponins containing a greater number of sugar chains are uncommon (Augustin et al., 2011). Monodesmosidic saponins are usually glycosylated at the *C*-3 hydroxyl group, while bidesmosidic saponins are typically glycosylated at the *C*-3 hydroxyl and *C*-28 carboxyl groups (for pentacyclic saponins) (Augustin et al., 2011; Erthmann et al., 2018; Augustin et al., 2012; Meesapyodsuk et al., 2007; Naoumkina et al., 2010; Shibuya et al., 2010). For example, QS-21, a mixture of predominantly two saponins derived from an extract from *Quillaja saponaria*, contains a branched trisaccharide (glucuronic acid, xylose and galactose) at the *C*-3 hydroxyl group and a linear tetrasaccharide (fucose, rhamnose, xylose, and apiose/xylose) at the *C*-28 carboxyl group (Figure 1.6) (Ragupathi et al., 2011). By contrast, mono- and bidesmosidic ginsenosides, which are tetracyclic, are typically glycosylated at the *C*-3, *C*-6 and/or *C*-20 hydroxyl groups (Wei et al., 2015; Wang et al., 2015; Yan et al., 2014).

Similar to P450s, UGTs are grouped into families and subfamilies based on amino acid sequence identity, with families and subfamilies typically sharing > 40 % and > 60 % sequence identity, respectively (Augustin et al., 2011). To date, more than 20 UGTs involved in triterpenoid biosynthesis have been identified across at least four families and seven subfamilies, many of which belong to the UGT73 family (Augustin

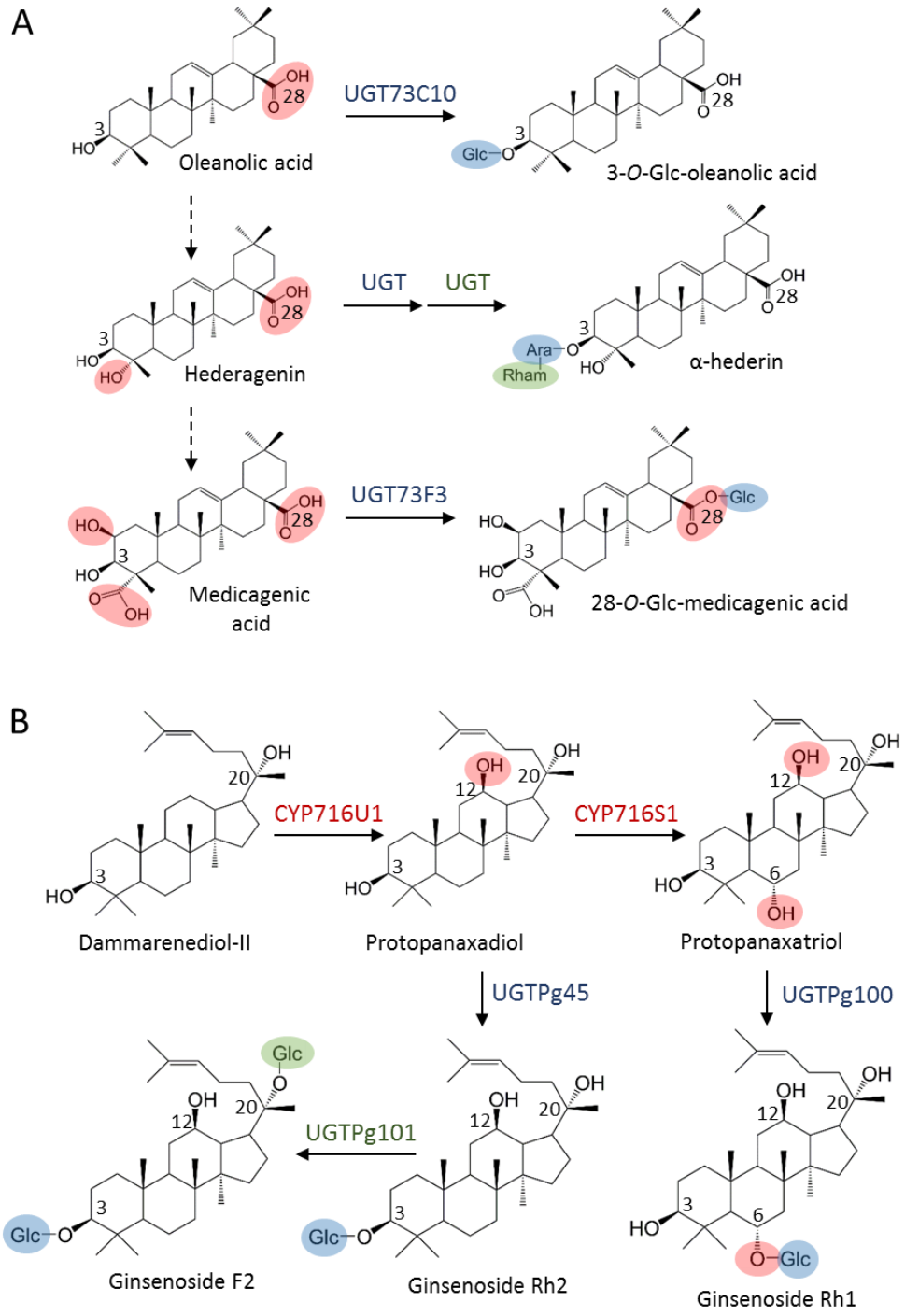


Figure 1.5. Glycosylation of triterpenoids produces diverse saponins. (A) Glycosylation of oleanolic acid and its oxygenated derivatives at the *C*-3 hydroxyl and *C*-28 carboxyl groups, and the generation of a disaccharide chain at *C*-3 through the sequential action of two UGTs. **(B)** Glycosylation of the dammarenediol-II derivatives protopanaxadiol and protopanaxatriol at *C*-3, *C*-6 and *C*-20 to produce ginsenosides. Dashed arrows represent multiple reactions. Moieties introduced by P450s are highlighted red, while sugars are labelled blue and green.

et al., 2012; Erthmann et al., 2018; Meesapyodsuk et al., 2007; Naoumkina et al., 2010; Achnine et al., 2005; Shibuya et al., 2010; Sayama et al., 2012; Wei et al., 2015; Wang et al., 2015; Yan et al., 2014). The identified UGTs predominantly glycosylate pentacyclic triterpenoids at the *C*-3 hydroxyl and *C*-28 carboxyl groups (Augustin et al., 2012; Erthmann et al., 2018; Meesapyodsuk et al., 2007; Naoumkina et al., 2010), tetracyclic dammarane-type triterpenoids at the *C*-3, *C*-6 and *C*-20 hydroxyl groups (Wei et al., 2015; Yan et al., 2014; Wang et al., 2015), and/or extend existing saponin sugar chains on both tetracyclic and pentacyclic saponins (Shibuya et al., 2010; Sayama et al., 2012; Wang et al., 2015).

1.1.3.3 Other tailoring reactions

In addition to oxygenations and glycosylations, numerous other modifications can also occur, including acylation, malonylation and methylation (Moses et al., 2014a). For example, *Avena strigosa* (oat) produces β-amyrin-derived saponins called avenacins, which are important for disease resistance (Papadopoulou et al., 1999). They are acylated at the *C*-21 hydroxyl by AsSCPL1 (*A. strigosa* serine carboxypeptidase-like acyltransferase 1), which transfers either an *N*-methyl anthranilate moiety (producing avenacins A-1 and B-1) or a benzoate moiety (producing avenacins A-2 and B-2) (Mugford et al., 2009). Mutants of *A. strigosa* lacking a functional AsSCPL1 accumulate avenacins lacking the acyl group (des-acyl avenacins) and are more susceptible to disease (Papadopoulou et al., 1999; Qi et al., 2004).

Meanwhile, QS-21 contains an acyl chain attached to the fucose moiety of its *C*-28 tetrasaccharide, and this acyl chain is further glycosylated with an arabinose moiety (Figure 1.6) (Ragupathi et al., 2011). The acyl chain may be important for the ability of QS-21 to act as an adjuvant, as it may stimulate the proliferation of T-killer cells (Ragupathi et al., 2011). Other examples include the saponin vaccaroside B from *Saponaria vaccaria*, which has 2-hydroxy-2-methylglutarate attached to its *C*-28 tetrasaccharide chain (Meesapyodsuk et al., 2007), and the triterpenoid cinnamoyl-lupeol, which has a 4-hydroxy-cinnamoyl moiety at the *C*-3 hydroxyl (Williams, 1999).

In summary, saponin biosynthesis involves four main types of enzymes: (i) OSCs, catalysing the cyclisation of 2,3-oxidosqualene to a triterpene scaffold, (ii) P450s that oxygenate the scaffold, (iii) UGTs that perform glycosylation reactions at various functional groups, and (iv) other tailoring enzymes, such as acyltransferases.

1.2 Properties and applications of triterpenoids

The huge structural diversity of triterpenoids results in diverse properties and bioactivities, with applications in numerous sectors (Moses et al., 2014a). While triterpenes are hydrophobic, typically containing a single hydroxyl group derived from the epoxide of 2,3-oxidosqualene, oxygenation of triterpenes by P450s increases their polarity (Moses et al., 2014a). Meanwhile, glycosylation introduces a large hydrophilic sugar moiety and the resulting saponin is therefore amphipathic (Moses et al., 2014a). As such, saponins can act as surfactants, foaming agents or even permeabilise cell membranes (Augustin et al., 2011; Moses et al., 2014a). The tailoring reactions thus have a major impact on the properties of the molecule and are essential for bioactivity (Moses et al., 2014a; Ragupathi et al., 2011; Nielsen et al., 2010; Geisler et al., 2013; Papadopoulou et al., 1999). Examples of triterpenoids and saponins with medically and commercially relevant properties are discussed (Figure 1.6).

The ginsenosides are a class of saponin produced by the glycosylation of dammarane-type triterpenoids, specifically protopanaxadiol and protopanaxatriol, which are derived from the oxygenation of the tetracyclic triterpene dammarenediol-II (Dai et al., 2013). They are produced by *Panax* spp. (ginseng), a traditional Chinese medicine, and have shown anti-cancer activities in clinical trials (Nag, 2012). Furthermore, ginsenoside compound K (Figure 1.6) is being investigated in clinical trials in China for the treatment of arthritis (Chen et al., 2017; Yang et al., 2015).

Various triterpenoids of the lupane class, which are derived from the triterpene lupeol, also display interesting bioactivities. Betulinic acid, which is the *C*-28 carboxyl derivative of lupeol, is a potential anti-cancer compound that has been investigated in clinical trials (Figure 1.6) (Moses et al., 2013). Meanwhile, bevirimat, a betulinic acid derivative that inhibits the replication of HIV type 1, was evaluated in a phase II clinical trial to treat HIV (Figure 1.6) (Smith et al., 2007). In contrast, cinnamoyl-lupeol, which is derived from lupeol by the addition of a cinnamoyl group, showed insecticidal activities against *Cylas formicarius* (sweet potato weevil), a pest of sweet potato (Williams, 1999).

The oleanane triterpenoids, which are derived from the triterpene β-amyrin, contain numerous molecules with diverse properties and bioactivities. Oleanolic acid (Figure 1.6), the *C*-28 carboxyl derivative of β-amyrin, has hepatoprotective activities and is used to treat viral hepatitis in China (Pollier and Goossens, 2012). Furthermore,

preliminary results from a clinical trial indicated that it has potential as a hypolipidemic drug (Lin et al., 2016). Oleanolic acid has also shown anti-cancer and anti-inflammatory activities (Pollier and Goossens, 2012), which lead to the development of a series of synthetic oleanolic acid derivatives with enhanced activities (the bardoxolone series) (Figure 1.6). The first derivative, bardoxolone, was evaluated in a phase I clinical trial with leukaemia patients (Tsao et al., 2010), while bardoxolone methyl underwent phase II and III clinical trials for patients with chronic kidney disease and type II diabetes, although it showed adverse effects (Moses et al., 2013). Meanwhile, bardoxolone imidazolide and bardoxolone ethylamide have the potential to treat osteoarthritis due to their anti-inflammatory activities and ability to induce chondrogenesis (Suh et al., 2012). Finally, the synthetic oleanolic acid derivative RTA 408 (Figure 1.6) is under clinical development by Reata Pharmaceuticals Inc., and is currently undergoing a phase II clinical trial for the treatment of the neurodegenerative disease Friedreich's ataxia (Strawser et al., 2017).

Numerous saponins of the oleanane class also display medically relevant bioactivities. For example, QS-21 is being investigated as an adjuvant in various vaccines including against HIV and malaria, and is commercially available in a veterinary vaccine against feline leukaemia virus (Moses et al., 2013; Ragupathi et al., 2011; Mahmoudi and Keshavarz, 2017; Sun et al., 2009). The biosynthesis of QS-21 from β-amyrin involves multiple oxygenations, glycosylations, and an acylation (Figure 1.6). The acyl chain is critical for adjuvant activity, as its removal results in loss of its immunostimulatory activity (Ragupathi et al., 2011). The sugar chains, the aldehyde group at *C*-23, and its amphipathic structure are also important for its adjuvant activity (Ragupathi et al., 2011).

Saponins derived from hederagenin (23-hydroxy-oleanolic acid) have been identified as having anti-feedant properties against plant pests, and may therefore have agricultural applications. Hederagenin cellobioside (3-*O*-[glucose-glucose]-hederagenin), α-hederin (3-*O*-[rhamnose-arabinose]-hederagenin), and 3-*O*-glucose-hederagenin (Figure 1.6) all acted as feeding deterrents against *Phyllotreta nemorum* (flea beetle) (Nielsen et al., 2010; Augustin et al., 2012). Hederagenin cellobioside also inhibited feeding by *Plutella xylostella* (diamondback moth) (Shinoda et al., 2002). Meanwhile, the corresponding saponins with oleanolic acid aglycones (oleanolic acid cellobioside and 3-*O*-glucose-oleanolic acid), which lack the *C*-23

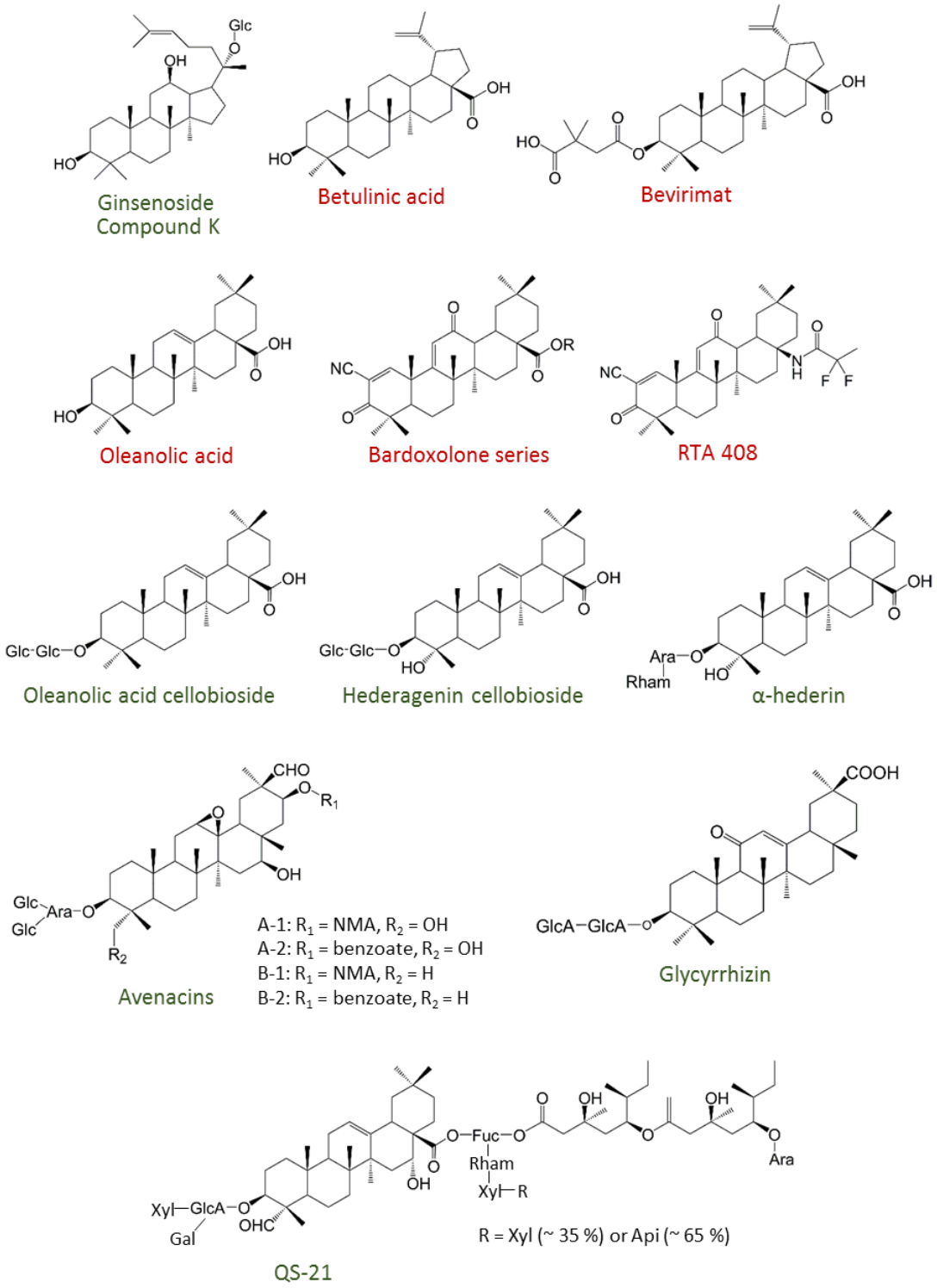


Figure 1.6. Triterpenoids and saponins with industrial potential. For the bardoxolone series, the R group is H (bardoxolone), methyl (bardoxolone methyl), imidazolide (bardoxolone imidazolide) or ethylamide (bardoxolone ethylamide). Triterpenoid aglycones and saponins are labelled in red and green, respectively. Glc: glucose; Ara: arabinose; Rham: rhamnose; GlcA: glucuronic acid; Gal: galactose; Xyl: xylose; Fuc: fucose; Api: apiose; NMA: *N*-methyl anthranilate

hydroxyl group, had much weaker anti-feedant activities, indicating that this moiety is important for activity (Nielsen et al., 2010; Augustin et al., 2012). Neither of the corresponding aglycones (hederagenin and oleanolic acid) inhibited feeding, demonstrating the importance of glycosylation for activity (Nielsen et al., 2010).

Similarly, the avenacins (Figure 1.6) are required for disease resistance (against fungal pathogens) in oat (*Avena spp.*) (Papadopoulou et al., 1999) and therefore also have potential applications in agriculture (Haralampidis et al., 2001). The avenacins have a β-amyrin backbone that is oxygenated at four to five positions (depending on the avenacin), contain a branched trisaccharide at the *C*-3 hydroxyl, and are acylated with either benzoate or *N*-methyl anthranilate at the *C*-21 hydroxyl (Figure 1.6) (Mugford et al., 2009). It was reported that the *C*-12,13 epoxide, which is introduced by CYP51H10, is important for their anti-fungal activities (Geisler et al., 2013). Furthermore, *Avena strigosa* mutants that don't accumulate acylated avenacins are more susceptible to fungal attack than wild type plants, indicating that acylation of the avenacins may also be important for disease resistance (Papadopoulou et al., 1999; Qi et al., 2004).

In addition to its ability to act as a feeding deterrent, α-hederin displayed anti-cancer activities in mouse models (Liu et al., 2014; Bang et al., 2005). However, it also exhibited a strong haemolytic activity, lysing red blood cells, that may preclude its use in the clinic (Gauthier et al., 2009; Tong et al., 2017). Indeed, numerous oleanane saponins have haemolytic activities (Moses et al., 2014a). This is due to their ability to permeabilise cell membranes, which can be attributed to the amphipathic structure of saponins (Augustin et al., 2011). However, not all saponins have this activity. For example, lupane-type saponins do not permeabilise cell membranes, and nor do some oleanane saponins (Moses et al., 2014a). Indeed, the oleanane saponins produced by *Medicago truncatula* can be classed as haemolytic or non-haemolytic, the former having a *C*-28 carboxyl group (being derived from oleanolic acid) and the latter having a *C*-24 hydroxyl group (being derived from 24-hydroxy-β-amyrin); the *C*-28 carboxyl may therefore be important for haemolytic activity (Seki et al., 2011; Moses et al., 2014a).

Hederacolchiside A, which is derived from α-hederin by the addition of a glucose moiety to the existing sugar chain, showed stronger antitumour activity than α-hederin in mouse models (Bang et al., 2005). This activity was even greater than the

commercially available anti-cancer drugs taxol and doxorubicin (Bang et al., 2005). However, as with α-hederin, it exhibits considerable haemolytic activity (Gauthier et al., 2009). By generating a series of synthetic derivatives of α-hederin and its aglycone hederagenin, Tong *et al.* (2017) identified a hederagenin derivative that showed strong cytotoxicity against human cancer cell lines and minimal haemolytic activity.

Finally, the oleanane saponin glycyrrhizin (Figure 1.6) is responsible for the sweetness of liquorice and is widely used as a sweetener (Osbourn et al., 2011), having a sweetness approximately 170 times greater than sucrose (Mizutani et al., 1994). In addition, glycyrrhizin has anti-viral activities, showing promise in clinical trials for the treatment of HIV/AIDs and is currently being used in Japan to treat chronic hepatitis (Fiore et al., 2008). Glycyrrhizin is derived from β-amyrin via two oxygenations and the sequential transfer of two glucuronic acid moieties to the *C*-3 hydroxyl, generating a disaccharide sugar chain (Seki et al., 2008, 2011; Xu et al., 2016).

1.3 Methods to produce triterpenoids and saponins

Because of their diverse applications there is great interest in the medical and commercial exploitation of triterpenoids and saponins. However, current approaches to triterpenoid production have considerable limitations, and are typically unable to provide triterpenoids in the high abundances and purities required for commercial exploitation (Moses et al., 2013). Methods of obtaining plant natural products such as triterpenoids include extraction from the native plant, chemical synthesis and biotechnological production, in which organisms (often microbes) are engineered to produce the compounds of interest (Moses et al., 2013; Osbourn et al., 2011; Kotopka et al., 2018). The latter approach has many advantages for the production of complex metabolites such as triterpenoids and saponins. These approaches are discussed below.

The traditional approach of obtaining triterpenoids and saponins (and many other plant natural products besides) is extraction from the native plant (Moses et al., 2013). For example, the triterpenoid oleanolic acid is currently obtained from olive (Pollier and Goossens, 2012). However, plants typically produce numerous structurally related triterpenoids and saponins in low abundance, and often under specific conditions (e.g. stress or pathogen attack), which complicates the isolation of individual molecules, leading to low yields and purities (Moses et al., 2013, 2014a).

Indeed, natural saponin extracts, such as QuilA from *Quillaja saponaria*, comprise a mixture of saponins instead of a single structure (Moses et al., 2014a). Furthermore, most such plants are not crop plants and are therefore difficult to cultivate. Yields can vary due to environmental fluctuations and the timescale for production is typically many months or years (Moses et al., 2013; Paddon and Keasling, 2014). As such, this method of production can be challenging and time-consuming, requiring either large areas of land for the cultivation of crops or the environmentally destructive process of extracting metabolites from wild plants.

A second approach is chemical synthesis, which is used for the production of various natural products (Nicolaou et al., 1998). However, this method is often unsuitable for the synthesis of complex molecules such as triterpenoids and saponins, whose biosynthesis involves the generation of multiple chiral centres and regio- and stereospecific oxygenations (Hoshino, 2017; Thimmappa et al., 2014). While the total chemical synthesis of several saponins has been reported, these involved numerous reaction steps and repeated purifications, during which much of the product was lost (Zhu et al., 2008; Cheng et al., 2009; Xiao and Yu, 2013; Deng et al., 1999; Kim et al., 2013).

An alternative method to produce plant natural products is the application of biotechnology to produce saponins in microorganisms, in which the relevant plant biosynthetic enzymes are heterologously expressed in a chosen microorganism. This approach has the potential to overcome many of the limitations of the above methods. Complex molecules with multiple chiral centres and regio- and stereospecific modifications can be produced relatively easily, as this approach utilises the plant enzymes that naturally catalyse such reactions. Furthermore, because heterologous pathways are free from their native regulation, expression can be increased or fine-tuned to maximise flux through the pathway, enabling high-level production of the chosen product (Kotopka et al., 2018; Moses et al., 2013). Model microorganisms such as *S. cerevisiae* and *Escherichia coli* are genetically tractable and well characterised, which enables extensive engineering of their native metabolism to redirect metabolic flux towards the heterologous pathway (Moses et al., 2013). In addition, the high growth rates of microorganisms reduces the production time compared with plants (Kotopka et al., 2018; Moses et al., 2013), and because microbial fermentation processes are performed in closed fermenters, they are not

affected by the environmental fluctuations that affect the yields of plant production systems. Purification is also often easier, as the engineered microorganisms accumulate fewer metabolites that are structurally similar to the product (Kotopka et al., 2018; Moses et al., 2013).

The biotechnological approach is exemplified by the semi-synthetic production of the sesquiterpene artemisinin (Paddon and Keasling, 2014). This antimalarial drug has traditionally been obtained through the farming of *Artemisia annua* (sweet wormwood), however, large fluctuations in the price and availability of artemisinin have caused difficulties in supplying the drug to the developing countries where it is needed (Paddon and Keasling, 2014). These fluctuations were attributed to the long (18 month) period required to produce and supply artemisinin, as well as fluctuations in weather and harvest (Paddon and Keasling, 2014). In the semi-synthetic approach, *S. cerevisiae* was engineered to produce artemisinic acid, the precursor to artemisinin, and through a combination of metabolic engineering and process development, titres of 25 g/L artemisinic acid were achieved (Paddon et al., 2013). Artemisinic acid could be purified from the culture with a yield of 93 % and a purity of 98 % (by weight), and then converted to artemisinin through a synthetic chemical approach (Paddon et al., 2013). Semi-synthetic artemisinin has since entered commercial production (Paddon and Keasling, 2014).

The yeast *S. cerevisiae* is a promising host for the industrial-scale production of triterpenoids and saponins. It is genetically tractable, fast growing, has a well-characterised metabolism and physiology (Kotopka et al., 2018), and numerous tools have been developed to aid engineering efforts (Lee et al., 2015; Apel et al., 2017; Curran et al., 2013; Ryan et al., 2014; Horwitz et al., 2015; Rajkumar et al., 2016; Blazeck et al., 2012; Redden and Alper, 2015; Curran et al., 2014). Furthermore, it is widely used in established industrial production processes (e.g. for bioethanol and artemisinin production (Mohd Azhar et al., 2017; Paddon et al., 2013; Paddon and Keasling, 2014)), and has been successfully used to produce a range of terpenoids at laboratory scale (Immethun et al., 2013). Yeast also produces 2,3-oxidosqualene, the first dedicated intermediate in triterpenoid biosynthesis (Moses et al., 2013), and is an effective host for the functional expression of plant P450s, which are typically localised to the ER and are extensively involved in triterpenoid biosynthesis (Kotopka et al., 2018; Thimmappa et al., 2014; Bak et al., 2011). Indeed, *S. cerevisiae* has been

successfully used for the laboratory-scale production of triterpenoids and saponins, as well as to characterise numerous enzymes involved in their biosynthesis (Kirby et al., 2008; Fukushima et al., 2011, 2013; Seki et al., 2011; Moses et al., 2014b; Czarnotta et al., 2017; Zhao et al., 2017; Zhuang et al., 2017a; Arendt et al., 2017; Zhao et al., 2018; Zhu et al., 2018; Fiallos-Jurado et al., 2016; Moses et al., 2015b, 2015a). Efforts to engineer yeast for the high-level production of triterpenoids and saponins are discussed in the following section.

1.4 Production of triterpenoids and saponins in *Saccharomyces cerevisiae*

Yeast engineering efforts have focused primarily on three classes of triterpenoids: (i) the oleananes, which are derived from the pentacyclic β-amyrin; (ii) the dammaranes, which are derived from the tetracyclic triterpene dammarenediol-II; and (iii) the lupanes, which are derived from the pentacyclic lupeol (Figure 1.3). While saponins of the dammarane and oleanane classes have been produced in yeast (Moses et al., 2014b; Arendt et al., 2017; Wang et al., 2015; Wei et al., 2015; Zhuang et al., 2017a), for the lupane class only aglycones have been produced to date (Huang et al., 2012; Moses et al., 2014b; Zhou et al., 2016; Czarnotta et al., 2017). Other types of triterpenoids, such as those belonging to the ursane class of α-amyrin derived triterpenoids, have also been produced to a more limited extent (Dong et al., 2018; Fukushima et al., 2011; Moses et al., 2014b).

1.4.1 Studies focusing on the production of oleananes

The first reported production of a triterpene in *S. cerevisiae* was by Kirby *et al.* (2008), who produced β-amyrin at a titre of 6 mg/L. This was achieved through the co-expression of the β-amyrin synthase (BAS) from *Artemisia annua* (*AaBAS*) and an N-terminally truncated variant of the yeast HMG-CoA reductase 1 gene (*tHMG1*), together with repression of the yeast lanosterol synthase gene (*ERG7*) using a methionine-repressible promoter (*pMET3*). By comparison, expression of *AaBAS* alone resulted in β-amyrin titres of 4 mg/L (Kirby et al., 2008). Yeast has two isozymes of HMG-CoA reductase (Hmg1p and Hmg2p), which catalyse the rate-limiting conversion of HMG-CoA to mevalonate in the MEV pathway (Figure 1.1), and it was previously shown that overexpression of *tHMG1* increased squalene production (Donald et al., 1997; Polakowski et al., 1998). Meanwhile, Erg7p competes with the BAS for the substrate 2,3-oxidosqualene, and its repression should therefore redirect

flux towards β-amyrin biosynthesis (Kirby et al., 2008). The additional overexpression of *ERG1*, the yeast squalene epoxidase that catalyses the rate-limiting conversion of squalene to 2,3-oxidosqualene (Figure 1.1), did not improve production further (Kirby et al., 2008).

Shortly thereafter, Seki *et al.* (2008) reported the production of 1.6 mg/L 11-oxo-β-amyrin, an intermediate in the biosynthesis of the saponin glycyrrhizin. This titre was achieved through the co-expression of *Lotus japonicus* BAS (*LjBAS*), a β-amyrin *C*-11 oxidase from *Glycyrrhiza uralensis* (*CYP88D6*) and a cytochrome P450 reductase from *L. japonicus* (*LjCPR1*). The CYP88D6 enzyme oxidises *C*-11 of β-amyrin to a ketone group, but small amounts (0.2 mg/L) of 11-hydroxy-β-amyrin were also observed, indicating that the reaction proceeds through a hydroxyl intermediate (Seki et al., 2008). In a follow-up to their study, Seki *et al.* (2011) additionally expressed *CYP72A154* from *G. uralensis*, which oxidises the *C*-30 methyl group of 11-oxo-β-amyrin to a carboxyl group. This generated glycyrrhetinic acid, the aglycone of glycyrrhizin, with titres of 15 µg/L (Seki et al., 2011). Derivatives of 11-oxo-β-amyrin with hydroxyl or aldehyde groups at *C*-30 (30-hydroxy-11-oxo-β-amyrin and glycyrrhetaldehyde, respectively) were also observed, indicating that the reaction occurs through hydroxyl and aldehyde intermediates. In addition, 30-hydroxy-β-amyrin was detected (indicating that CYP72A154 can also oxidise β-amyrin), as well as an isomer of glycyrrhetinic acid, considered to most likely have a *C*-29 carboxyl group instead of a *C*-30 carboxyl (Seki et al., 2011).

Fukushima *et al.* (2013) used a combinatorial approach to produce a variety of *M. truncatula* oleanane triterpenoids in yeast. This involved the co-expression of different pairs of P450s with *LjBAS* and *LjCPR*. Using this approach, they produced soyasapogenol B and gypsogenic acid, which are intermediates in the biosynthesis of the non-haemolytic (soyasaponin) and haemolytic classes of *M. truncatula* saponins, respectively (Fukushima et al., 2013). Furthermore, by co-expressing P450s involved in soyasaponin and haemolytic saponin biosynthesis, they were able to produce rare triterpenoids that do not accumulate in considerable abundance in *M. truncatula*. For example, co-expression of *M. truncatula CYP716A12* (encoding a β-amyrin *C*-28 oxidase involved in haemolytic saponin biosynthesis) and *CYP93E2* (encoding a *C*-24 hydroxylase involved in soyasaponin biosynthesis) produced 4-*epi*-hederagenin, which has not been reported in *M. truncatula* (Fukushima et al., 2013).

This report highlighted another advantage of microbial production systems: the ability to produce novel or rare triterpenoids through the co-expression of new combinations of tailoring enzymes such as P450s, an approach known as combinatorial biosynthesis (Moses et al., 2013).

The studies discussed so far reported either low titres (up to 6 mg/L) or no titres at all. By contrast, industrial-scale production would require titres of at least the gram per litre scale (for example, the 25 g/L artemisinic acid reported by Paddon *et al.* (2013)). Through a combination of metabolic engineering and process optimisation, Moses *et al.* (2014b) achieved β-amyrin titres of ~ 50 mg/L in shake flasks. Initially, BASs from *M. truncatula* and *Glycyrrhiza glabra* (*MtBAS* and *GgBAS*) were compared in a strain engineered to enhance flux towards triterpene biosynthesis, giving β-amyrin titres of 19 and 36 mg/L, respectively (Moses et al., 2014b). This strain overexpressed *tHMG1* and had the *ERG7* gene repressed using the *pMET3* promoter, the same strategy used by Kirby *et al.* (2008). In addition, expression of the *A. thaliana* lupeol synthase (*AtLUS1*) gave lupeol titres of 46 mg/L. These titres were considerably higher than those of Kirby *et al.* (2008), who achieved only 6 mg/L β-amyrin. This could be explained by the different BASs used, or due to differences in culturing and metabolite extractions. Notably, Moses *et al.* induced *ERG7* repression (by addition of 1 mM methionine) at 24 h, while Kirby *et al.* added methionine only at 43 h. Furthermore, it is possible that MtBAS and GgBAS have greater activities than AaBAS, which was used by Kirby *et al.* (2008).

Subsequently, co-expression of *GgBAS* or an α-amyrin synthase from *Centella asiatica* (*CaDDS*) with *CYP716Y1* (a 16α-hydroxylase) and the *A. thaliana* CPR *AtATR1* produced 16α-hydroxy-β-amyrin or 16α-hydroxy-α-amyrin, respectively (Moses et al., 2014b). By contrast, co-expression of *AtLUS1* with *CYP716Y1* and *AtATR1* produced no additional product, indicating that CYP716Y1 could not oxidise lupeol (Moses et al., 2014b). Because the ratio between the P450 and CPR has been reported to be important for activity (Paddon et al., 2013), different ratios were tested by expressing *CYP716Y1* from a high copy plasmid and *AtATR1* from an integrated, low or high copy plasmid (Moses et al., 2014b). The integrated gene resulted in very low levels of 16α-hydroxy-α-amyrin, while expression from the high and low copy plasmids gave considerably higher titres (Moses et al., 2014b). The addition of methyl-β-cyclodextrin (MβCD) to the culture medium further increased β-amyrin and 16α-

hydroxyl-β-amyrin levels, with β-amyrin titres of ~ 50 mg/L reported (Moses et al., 2014b). MβCD is a toroidal shaped molecule with a hydrophobic interior and hydrophilic exterior, which may allow it to encapsulate and solubilise the hydrophobic triterpenoids, sequestering them into the growth medium (Moses et al., 2014b). This could minimise growth defects or feedback inhibition caused by the intracellular accumulation of triterpenoids. Finally, co-expression of *GgBAS*, *CYP716A12*, *CYP716Y1* and *AtATR1* produced echinocystic acid (16α-hydroxy-oleanolic acid) (Moses et al., 2014b). The additional expression of *UGT73C11*, encoding a glucosyltransferase from *Barbarea vulgaris* with specificity for the *C*-3 hydroxyl group, produced the saponin 3-*O*-glucose-echinocystic acid, which was the first reported production of an oleanane saponin in yeast (Moses et al., 2014b).

By integrating *GgBAS* under the control of a constitutive promoter, Dai *et al.* (2014) achieved β-amyrin titres of 78 mg/L, approximately ~ 2-fold higher than the 36 mg/L titre reported by Moses *et al.* (2014b) when expressing *GgBAS*. Interestingly, whereas the strain used by Moses *et al.* overexpressed *tHMG1* and had *ERG7* repressed, Dai *et al.* used a "wild-type" strain that was not engineered to boost triterpenoid production (Dai et al., 2014; Moses et al., 2014b). Notably, Moses *et al.* (2014b) expressed *GgBAS* from the galactose-inducible *GAL1* promoter on a high copy plasmid, which required the use of synthetic dropout media lacking glucose (which represses the *GAL1* promoter). By using a constitutive promoter and integrating *GgBAS*, Dai *et al.* (2014) were able to grow the cells in rich media containing glucose, which is the preferred carbon source of *S. cerevisiae* (Timson, 2007). Furthermore, the culture period was seven days, whereas Moses *et al.* (2014b) used a three day culture period. Together, these differences could explain the observed differences in β-amyrin titres.

Dai *et al.* (2014) also compared a BAS from *Panax ginseng* (*PgBAS1*) with *GgBAS*, but *PgBAS1* gave far lower titres (2 mg/L compared with 78 mg/L). Using *GgBAS*, they further enhanced β-amyrin production to 107 mg/L by overexpressing *tHMG1*, *ERG9* and *ERG1* from constitutive promoters. The *ERG9* gene encodes yeast squalene synthase, which catalyses the joining of two molecules of FPP to produce squalene (Figure 1.1). Overexpression of these genes should thus increase flux towards β-amyrin biosynthesis. Finally, the additional expression of *CYP716A12*, *AtATR1* and an additional copy of *GgBAS* allowed the production of 71 mg/L oleanolic acid (Dai *et al.* 2014). Beta-amyrin also accumulated at 88.6 mg/L in this strain,

indicating that the conversion of β-amyrin to oleanolic acid was rate-limiting (Dai *et al.* 2014).

Meanwhile, Zhang *et al.* (2015) integrated and expressed *GgBAS* and the *Candida albicans* squalene epoxidase (*CaERG1*) from constitutive promoters, giving β-amyrin titres of 24.5 mg/L. The additional expression of *E. coli* IPP-DMAPP isomerase (*EcIDI*), yeast FPP synthase (*ERG20*) and *ERG9* increased titres to 36.5 mg/L. Next, the sterol regulatory element, to which the transcriptional activator Upc2p binds and regulates the sterol pathway, was inserted into all promoters that controlled expression of the overexpressed genes. This increased β-amyrin titres to 49 mg/L. Finally, ethanol fed-batch fermentation, in which cells were grown in a 5 L fermenter and ethanol was fed every 12 h, produced 138 mg/L β-amyrin after 6 days (Zhang et al., 2015).

More recently, Zhao *et al.* (2018) reported oleanolic acid titres of 186 mg/L in shake flasks and 607 mg/L in a fermenter. To optimise the conversion of β-amyrin to oleanolic acid, they initially compared four CPRs: AtATR1, LjCPR, MtCPR (from *M. truncatula*) and GuCPR (from *G. uralensis*). The interaction between the heterologous P450 and CPR is important for activity (Zhao et al., 2016; Zangar et al., 2004), and an arbitrary choice of CPR may not result in optimal conversion of β-amyrin into oleanolic acid in yeast. The CPRs were co-expressed with *GgBAS* and *CYP716A12* on a high copy plasmid, controlled by the galactose-inducible *GAL1* and *GAL10* promoters, and cells were grown with galactose as the carbon source (Zhao et al., 2018). Expression of *MtCPR* gave the highest oleanolic acid titre (9 mg/L); by comparison, the least productive strain expressed *AtATR1* and produced only 2.5 mg/L oleanolic acid. Next, the pathway encoding MtCPR was integrated and the *GAL1* gene, which is responsible for the metabolism of galactose, was knocked out (Zhao et al., 2018). This should prevent galactose from being utilised as a carbon source, ensuring that galactose levels remain high and continue to induce expression of the oleanolic acid pathway throughout the fermentation. Cells were grown with both glucose and galactose, which served as the carbon source and inducer, respectively, and the maximum oleanolic acid titre reported was 41.3 mg/L (Zhao et al., 2018).

Interestingly, the additional knock out of the *GAL80* gene further improved oleanolic acid titres to 70.3 mg/L (Zhao et al., 2018). In the absence of galactose, Gal80p binds and inhibits Gal4p, the transcriptional activator of the *GAL* promoters (Timson, 2007).

When galactose is present, Gal3p prevents Gal80p from inhibiting Gal4p, leading to the activation of the *GAL* promoters (Timson, 2007). Thus, a *GAL80* knockout removes the need for galactose to induce expression of the *GAL* promoters. Given that galactose was provided in the medium, it is surprising that the knockout of *GAL80* improved titres, as the promoters should have been activated by the galactose. A possible explanation is that the amount of galactose used did not allow for full inhibition of Gal80p (note that galactose should not be depleted during the fermentation due to the *GAL1* knockout, which prevents metabolism of galactose). They next additionally overexpressed *tHMG1*, *ERG9* and *ERG1*, which increased oleanolic acid titres > 2.5-fold to 186 mg/L. Finally, batch fermentation in a 5 L fermenter enabled the production of 607 mg/L oleanolic acid, which is the highest reported titre to date (Zhao et al., 2018). Use of glucose as the carbon source probably limited expression of the oleanolic acid pathway, as glucose represses the *GAL* promoters through a mechanism mediated by the transcriptional repressor Mig1p (Timson, 2007). Using a different carbon source or promoters that are not repressed by glucose might therefore further boost production.

Recently, the production of glycyrrhetinic acid at titres of 19 mg/L was reported (Zhu et al., 2018). This was ~1300-fold higher than the 15 µg/L reported by Seki *et al.* (2011). Initially, a glycyrrhetinic acid pathway (*GgBAS*, *G. uralensis C*-11 oxidase *CYP88D6*, *G. uralensis C*-30 oxidase *CYP72A154*, and *AtATR1*) was expressed in a strain overexpressing *EcIDI*, *ERG20*, *ERG9*, *CaERG1* (as used by Zhang *et al.* (2015), see above). This resulted in the production of low amounts of glycyrrhetinic acid (33.7 µg/L), but much greater amounts of its precursor 11-oxo-β-amyrin (7.5 mg/L). Furthermore, β-amyrin accumulated to relatively high levels (14 mg/L), indicating that flux through the P450-catalysed oxygenation steps was limiting. To address this, they compared CYP88D6 with another *C*-11 oxidase from *G. uralensis* (Uni25647) for 11-oxo-β-amyrin production, finding that using the latter enzyme doubled titres. Six CPRs were then compared for glycyrrhetinic acid production in a strain expressing *Uni25647* and *CYP72A154*. The CPRs were MTR2 and MTR3 (*M. truncatula*), ATR1 and ATR2 (*A. thaliana*), LjCPR1 (*L. japonicus*), and GuCPR1 (*G. uralensis*). The latter gave the highest glycyrrhetinic acid (517 µg/L) and 11-oxo-β-amyrin (15.3 mg/L) titres yet (Zhu et al., 2018). Two *C*-30 oxidases (CYP72A154 and *M. truncatula* CYP72A63), which convert 11-oxo-β-amyrin to glycyrrhetinic acid, were then compared for glycyrrhetinic acid production in the strain expressing *Uni25647*

and *GuCPR1*. Using *CYP72A63* gave much greater titres, increasing glycyrrhetinic acid production ~ 14-fold from 517 µg/L to 7.4 mg/L (Zhu et al., 2018). Finally, ethanol fed-batch fermentation in a 5 L fermenter further increased glycyrrhetinic acid production to 19 mg/L (Zhu et al., 2018).

The comparison of CPR and P450s performed by Zhao *et al.* (2018) and Zhu *et al.* (2018) highlight the benefit of comparing homologues of biosynthetic enzymes to boost triterpenoid production. Indeed, Moses *et al.* (2014c) previously compared eight P450s for the CYP93E subfamily for the *C*-24 hydroxylation of β-amyrin, the first committed step in the biosynthesis of soyasaponins. The P450s showed a range of activities, accumulating varying amounts of 24-hydroxy-β-amyrin, the side-product 24-carboxy-β-amyrin, and the substrate β-amyrin (Moses et al., 2014c). In particular, whereas CYP93E2 had poor activity, accumulating predominantly β-amyrin and very little 24-hydroxy-β-amyrin, CYP93E9 produced 52-fold more 24-hydroxy-β-amyrin and accumulated ~ 4-fold less β-amyrin substrate (Moses et al., 2014c). Thus, comparison of homologues is an effective strategy to relieve the commonly observed bottlenecks at the P450-catalysed oxygenations of triterpenoids (Moses et al., 2014c; Zhu et al., 2018; Zhao et al., 2018).

Finally, Arendt *et al.* (2017) reported that knocking out the yeast *PAH1* gene considerably boosts the production of various triterpenoids and saponins. The *PAH1* gene encodes phosphatidic acid phosphatase, and its knockout was previously shown to cause proliferation of the ER (Santos-Rosa et al., 2005). Because numerous enzymes involved in the biosynthesis of 2,3-oxidosqualene, as well as the plant OSCs, P450s and CPRs involved in triterpenoid biosynthesis, are believed to be localised to the ER, such a proliferation could increase the intracellular concentrations of these proteins and boost productivity (Arendt et al., 2017). Indeed, Arendt *et al.* (2017) showed that the ER was greatly expanded and that levels of GgBAS and especially CYP716A12 proteins were enhanced compared with wild-type. Notably, the *PAH1* knockout increased titres of β-amyrin and medicagenic acid, which is derived from β-amyrin through the action of three P450s, ~ 8- and 6-fold (Arendt et al., 2017). In addition, production of the saponin 28-*O*-glucose-medicagenic acid was increased by an even greater degree (~ 16-fold) (Arendt et al., 2017). This strain also produced substantially higher amounts of saponin glucosides of bayogenin, hederagenin and gypsogenic acid (whose aglycones are intermediates in the

biosynthesis of medicagenic acid). This study therefore brings the total number of oleanane saponins produced in yeast to five (including the 3-*O*-glucose-echinocystic acid produced by Moses *et al.* (2014b)).

1.4.2 Studies focusing on the production of dammaranes

Much work has also been done in the production of dammarane-type triterpenoids, where considerably higher titres (up to 4.25 g/L) have been reported than for oleanane-type triterpenoids (Dai et al., 2013; Zhao et al., 2016, 2017). Notably, numerous studies have reported the production of ginsenosides, which are *Panax* spp. saponins derived from the dammarane triterpenoids protopanaxadiol and protopanaxatriol, with titres as high as 292 mg/L (Wei et al., 2015; Wang et al., 2015; Zhuang et al., 2017a). By contrast, no titres have been reported for oleanane saponins to date (Moses et al., 2014b; Arendt et al., 2017).

Dai *et al.* (2013) reported protopanaxadiol titres of 1189 mg/L. They initially integrated a protopanaxadiol biosynthetic pathway encoding *P. ginseng* dammarenediol-II synthase (*PgDDS*), *P. ginseng CYP716U1* and the *A. thaliana* CPR *AtATR1*. PgDDS cyclises 2,3-oxidosqualene to dammarenediol-II, while CYP716U1 catalyses the *C*-12 hydroxylation of dammarenediol-II to produce protopanaxadiol. This initial strain gave protopanaxadiol titres of 0.5 mg/g dry cell weight (DCW) (Dai et al., 2013). Subsequently, they overexpressed *tHMG1*, *ERG20*, *ERG9* and *ERG1*, which increased protopanaxadiol titres 17-fold to 8.45 mg/g DCW, with 10.96 mg/g DCW of dammarenediol-II accumulating. The overexpression of additional copies of *tHMG1*, *AtATR1* and *CYP716U1* shifted the product ratio towards protopanaxadiol, with 13.1 mg/g DCW protopanaxadiol and 8.2. mg/g DCW dammarenediol-II (Dai et al., 2013).

Glucose fed-batch fermentation of the optimised strain produced 812 mg/L protopanaxadiol (Dai et al., 2013). However, 776 mg/L dammarenediol-II still accumulated, indicating that its conversion to protopanaxadiol was limiting. To further improve production, Dai *et al.* used a two-phase fermentation method where an organic solvent was added to the culture to extract triterpenoids from the culture throughout the fermentation run. This is conceptually similar to the use of cyclodextrins by Moses *et al.* (2014b) to sequester triterpenoids in the culture medium, and has been successfully utilised for the production of artemisinic acid in yeast (Ro et al., 2006). Using methyl oleate as the solvent, titres of 1189 mg/L protopanaxadiol were achieved (Dai et al., 2013). However, in contrast with the previous single-phase

fermentation, dammarenediol-II was produced in greater amounts than protopanaxadiol (at titres of 1548 mg/L) (Dai et al., 2013). It is possible that extraction of dammarenediol-II by the organic solvent prevented it from being converted to protopanaxadiol, thus shifting the product ratio.

Zhao *et al.* (2016) reported slightly improved protopanaxadiol titres than the above study, but in contrast were able to achieve almost complete conversion of dammarenediol-II into protopanaxadiol. They initially followed a similar strategy to Dai *et al.* (2013), integrating the protopanaxadiol pathway, *tHMG1*, *ERG20*, *ERG9* and *ERG1*, which gave titres of 155.3 mg/L protopanaxadiol and 96.5 mg/L dammarenediol-II (Zhao et al., 2016). However, they subsequently sought to improve the conversion of dammarenediol-II into protopanaxadiol through fusion of the P450 and CPR, which may facilitate electron transfer from CPR to P450 and thus improve catalytic activity. Removal of the N-terminal membrane anchor of AtATR1 and fusion to the C-terminus of CYP716U1 greatly improved protopanaxadiol titres, which reached 265.7 mg/L, with only 8.7 mg/L dammarenediol-II accumulating (Dai et al., 2013). Thus, the fusion strategy allowed almost complete conversion of dammarenediol-II to protopanaxadiol. Finally, glucose fed-batch fermentation gave a protopanaxadiol titre of 1437 mg/L, with only 19.7 mg/L dammarenediol-II accumulating. Similar to the results of Dai *et al.* (2013), two-phase fed-batch fermentation using dodecane as the solvent resulted in a dramatic shift in product ratios, with the strain now accumulating mostly dammarenediol-II (524 mg/L) compared to 312.0 mg/L protopanaxadiol (Zhao et al., 2016). Thus, use of an organic overlay limits conversion of dammarenediol-II to protopanaxadiol, presumably by sequestration of dammarenediol-II that prevents it from being oxygenated by the P450.

The highest titre for a triterpenoid produced in yeast corresponds to 4.25 g/L protopanaxadiol, reported by Zhao *et al.* (2017). They reasoned that the generation of reactive oxygen species (ROS) and ethanol stress limited the productivity of the strain reported by Zhao *et al.* (2016). Poor coupling between the P450 and CPR can generate ROS through electron leakage (Zangar et al., 2004; Renault et al., 2014), and the ethanol produced as a waste product during fermentation can cause stress and even exacerbate the generation of ROS (Zhao et al., 2017). Indeed, the strain expressing the CYP716U1-AtATR1 fusion protein had significantly higher ROS levels

than a strain lacking this fusion protein (Zhao et al., 2017). Therefore, to increase tolerance towards ROS and ethanol, Zhao *et al.* (2017) overexpressed the yeast genes *YBP1* and *SSD1*. The former gene is essential for the activity of Yap1p, a transcriptional activator that mediates the yeast oxidative stress response. Meanwhile, *SSD1* is a translational repressor involved in cell wall maintenance, and enhancing cell wall integrity may improve tolerance to ethanol (Zhao et al., 2017). The resulting strain produced 1255 mg/L protopanaxadiol. Furthermore, by adding an ethanol feed to serve as an additional carbon source, titres were further improved to 1806 mg/L. Finally, a two-stage fermentation process, in which cells were grown in a fermenter with glucose, followed by the addition of ethanol to serve as the main carbon source once glucose was depleted, increased protopanaxadiol titres to 4.25 g/L (Zhao et al., 2017).

Wang *et al.* 2015 reported the production of the ginsenosides Rh2 and Rg3. The former is derived from protopanaxadiol through glycosylation at the *C*-3 hydroxyl group, while Rg3 is produced by the addition of a second glucose to the existing sugar chain of Rh2. They initially expressed a protopanaxadiol pathway (*PgDDS*, *CYP716U1* and *AtATR2*) together with overexpression of two copies of *tHMG1*, *ERG20*, *ERG9* and a squalene epoxidase from *P. ginseng* (*PgSQE1*). This strain produced 8.29 µmol/g DCW protopanaxadiol and accumulated minimal amounts of dammarenediol-II (0.14 µmol/g). This contrasts with the strains reported by Dai *et al.* (2013), which accumulated substantial amounts of dammarenediol-II. A possible explanation is the difference in CPR used (Dai et al., 2013; Wang et al., 2015). Indeed, it was previously reported that AtATR2 has a greater electron transfer activity than AtATR1 *in vitro* (Schückel et al., 2012), while the CPR comparison performed by Zhu *et al.* (2018) (see previous section) found that using AtATR2 gave higher glycyrrhetinic acid titres than AtATR1. Next, the additional expression of *UGTPg45*, which transfers a glucose to the *C*-3 hydroxyl of protopanaxadiol, produced ginsenoside Rh2 (3-*O*-glucose-protopanaxadiol) with a titre of 1.45 µmol/g (Wang et al., 2015). Finally, by co-expressing *UGTPg45* with *UGTPg29*, which transfers a glucose to the existing glucose of Rh2, ginsenoside Rg3 (3-*O*-[glucose-glucose]-protopanaxadiol) was produced with a titre of 3.49 µmol/g (Wang et al., 2015).

The production of ginsenosides derived from protopanaxatriol, which is produced by the *C*-6 hydroxylation of protopanaxadiol, have also been reported. Wei *et al.* (2015)

produced ginsenosides Rh1 and F1, which are derived from the glucosylation of protopanaxatriol at the *C*-6 and *C*-20 hydroxyl groups, respectively. Using the protopanaxadiol-producing strain reported by Wang *et al.* (2015), they additionally expressed *CYP716S1* (encoding a protopanaxadiol *C*-6 hydroxylase), an additional CPR from *P. ginseng* (*PgCPR1*), and either *UGTPg100* or *UGTPg1* (Wei et al., 2015). The first UGT transfers a glucose to the *C*-6 hydroxyl of protopanaxatriol, forming ginsenoside Rh1, while UGTPg1 transfers glucose to the *C*-20 hydroxyl to produce ginsenoside F1. Titres of 92.8 mg/L ginsenoside Rh1 and 42.1 mg/L ginsenoside F1 were achieved (Wei et al., 2015).

Finally, Zhuang *et al.* (2017) reported the production of 292 mg/L ginsenoside Rh2, the highest reported production of a saponin in yeast to date. They initially tested UGT51, a yeast UGT that glucosylates ergosterol, for activity against protopanaxadiol and protopanaxatriol *in vitro*. It had weak activity against these substrates, and the glucosylated protopanaxadiol was confirmed as 3-*O*-glucose-protopanaxadiol (ginsenoside Rh2) by NMR (Zhuang et al., 2017a). Subsequently, they engineered UGT51 for enhanced activity against protopanaxadiol using an iterative saturation mutagenesis method, with the best mutant having 610-fold higher activity against protopanaxadiol. Expression of this mutant in a yeast strain engineered for protopanaxadiol production resulted in the production of 6.1 mg/L ginsenoside Rh2; by comparison, only 0.05 mg/L Rh2 was produced when expressing the wild-type enzyme (Zhuang et al., 2017a). Deletion of the yeast *EGH1* gene, which encodes a glucosidase that was shown to hydrolyse ginsenoside Rh2 *in vitro*, doubled Rh2 titres to 12.2 mg/L, and the expression of three additional copies of the mutant *UGT51* further increased titres to 22.2 mg/L (Zhuang et al., 2017a). Because UGT51 requires UDP-glucose as a donor to glycosylate its substrate, the yeast UDP-glucose biosynthetic genes *PGM1* and *UGP1* were overexpressed, resulting in Rh2 titres of 36.7 mg/L (Zhuang et al., 2017a). Finally, the additional expression of a CYP716U1-AtATR2 fusion protein, based on the CYP716U1-AtATR1 fusion protein of Zhao *et al.* (2016), increased Rh2 titres to 45 mg/L. Glucose fed-batch fermentation of this strain gave titres of 292 mg/L ginsenoside Rh2 (Zhuang et al., 2017a).

1.4.3 Studies focusing on the production of lupanes

In the biotechnological production of lupane-type triterpenoids most studies have focused on betulinic acid, a potential anti-cancer compound that is the *C*-28 carboxy

derivative of lupeol (see section 1.2). The oxygenation reaction appears to proceed through hydroxyl (betulin) and aldehyde (betulinic aldehyde) intermediates (Fukushima et al., 2011; Huang et al., 2012). To date, no studies have reported the production of lupane-type saponins in yeast.

An initial study by Huang *et al.* (2012) produced 0.1 mg/L betulinic acid in yeast through the co-expression of a *Catharanthus roseus* multifunctional OSC (*CrAS*), the *C. roseus C*-28 oxidase *CYP716AL1*, and *AtATR1* (Huang et al., 2012). Zhou *et al.* (2016) later increased betulinic acid titres to ~ 0.6 mg/L. They initially compared three *C*-28 oxidases for the oxygenation of lupeol to betulinic acid (with co-expression of the *A. thaliana* lupeol synthase *AtLUP1* and *AtATR1*): CYP716AL1, CYP716A15 from *Vitis vinifera*, and CYP716A80 from *Barbarea vulgaris*. The latter enzyme resulted in the highest amounts of betulinic acid, with less accumulation of the hydroxyl intermediate betulin and the substrate lupeol compared with the other enzymes (Zhou et al., 2016). Finally, because the enzymes were expressed from the galactose-inducible *GAL* promoters, they knocked out the *GAL80* gene, which removes the requirement for galactose to induce *GAL* promoter expression. This gave titres of 0.6 mg/L betulinic acid when grown in galactose, which were ~ 2-fold higher than the wild-type strain (Zhou et al., 2016); improved triterpenoid titres when growing a *gal80* strain in galactose were also observed by Zhao *et al.* (2018) (see section 1.4.1).

The highest reported production of betulinic acid was achieved by Czarnotta *et al.* (2017), who produced 182 mg/L by optimising process conditions. They initially generated a strain expressing a betulinic acid pathway (*AtLUP1*, *CYP716AL1* and *AtATR2*) and the *ERG1* and *tHMG1* genes. This strain was grown in shake flasks with 50 g/L glucose, and glucose and ethanol levels were monitored. As expected, glucose was initially consumed and ethanol accumulated, followed by consumption of ethanol after glucose had depleted. Triterpenoids only accumulated during the ethanol consumption phase, and titres of betulinic acid and betulin reached 28 mg/L and 91 mg/L, respectively (Czarnotta et al., 2017). To increase triterpenoid production, an ethanol pulsing strategy to supply excess ethanol was investigated. Here, ethanol was added to the medium in response to high levels of dissolved oxygen, which are indicative of ethanol depletion. This approach increased betulinic acid titres to 57 mg/L, while betulin reached 167 mg/L (Zhou et al., 2016).

A further increase in production was achieved through the use of resting cell fermentations (Zhou et al., 2016). Cells were first grown to a high biomass with glucose as the carbon source, where triterpenoids were not expected to accumulate. Subsequently, the cells were transferred to fresh media lacking nitrogen, and the ethanol pulse feed was applied. The absence of nitrogen prevents cell growth, with the aim of redirecting metabolic flux from biomass generation to triterpenoid production. This strategy substantially increased triterpenoid production, with titres of betulinic acid and betulin reaching 182 mg/L and 464 mg/L, respectively (Zhou et al., 2016). It is notable that betulin accumulated to much higher levels than betulinic acid. As betulin is an intermediate in the oxygenation of lupeol to betulinic acid, this step is rate-limiting for betulinic acid production and requires optimisation. The strategies described to enhance the production of oxygenated oleanane- and dammarane-type triterpenoids, such as comparing different CPRs and P450 homologues, or generating P450-CPR fusions (Moses et al., 2014c; Zhu et al., 2018; Zhao et al., 2018, 2016; Zhuang et al., 2017a), show promise to alleviate this bottleneck in betulinic acid production.

1.5 Scope of this thesis

Numerous studies have engineered *S. cerevisiae* for the production of triterpenoids and saponins (see section 1.4). However, while high titres were achieved in some cases (notably for the dammarane-type triterpenoids), in many cases triterpenoid productivities remained low. In addition, a very limited range of oleanane saponins have been produced in yeast to date (Moses et al., 2014b; Arendt et al., 2017), despite their promising applications across numerous sectors. These saponins were relatively simple in structure, being monodesmosides and containing only a single sugar (Moses et al., 2014b; Arendt et al., 2017). By contrast, bioactive saponins, such as QS-21, can possess many sugars and multiple sugar chains (Ragupathi et al., 2011; Nielsen et al., 2010; Bang et al., 2005; Fiore et al., 2008). Thus, there is a need to both improve triterpenoid titres and to demonstrate the ability to produce more complex saponins in yeast.

The studies discussed in section 1.4 highlight the benefit of comparing enzyme variants to improve flux through a particular step of a biosynthetic pathway. Moses *et al.* (2014b) compared GgBAS and MtBAS for the production of β-amyrin in yeast, finding that GgBAS gave ~ 2-fold higher β-amyrin titres, while Dai *et al.* (2014) found

that expression of GgBAS gave a ~ 50-fold improvement in β-amyrin titres compared with PgBAS1. Similarly, the comparison of CYP93E enzymes performed by Moses *et al.* (2014c) found a diverse array of productivities, with some homologues converting very little of the substrate β-amyrin into 24-hydroxy-β-amyrin, and others having considerably higher catalytic activities. Finally, Zhao *et al.* (2018) and Zhu *et al.* (2018) compared various CPRs to improve the conversion of β-amyrin into oleanolic acid or glycyrrhetinic acid, respectively. Again, this strategy considerably boosted triterpenoid titres.

The aim of this thesis is to engineer *S. cerevisiae* to produce oleanane-type triterpenoids and saponins. The oleananes have numerous medically and commercially relevant properties (see section 1.2) that make them excellent targets for production in yeast. In particular, this study focuses on the production of β-amyrin and oleanolic acid, and saponins derived from them. The production of β-amyrin is the first committed step in the biosynthesis of oleananes (Thimmappa et al., 2014), which contain numerous bioactive compounds, while oleanolic acid is a key oleanane-type triterpenoid that is the precursor to many of the bioactive oleanane triterpenoids and saponins (Ragupathi et al., 2011; Nielsen et al., 2010; Bang et al., 2005; Fiore et al., 2008; Augustin et al., 2012; Lin et al., 2016).

In Chapter 3, a gas chromatography-mass spectrometry (GC-MS) method to monitor and quantify triterpenoid production in yeast is presented.

In Chapter 4, β-amyrin synthase (BAS) homologues are compared for the production of β-amyrin in yeast, with the aim of identifying a homologue that enables enhanced β-amyrin accumulation. Similarly, in Chapter 5 various cytochromes P450 of the CYP716A subfamily are compared for their ability to oxidise β-amyrin to oleanolic acid. As this oxygenation proceeds through two intermediates (erythrodiol and oleanolic aldehyde), product ratios of the P450s are also compared.

Finally, Chapter 6 explores the production of saponins derived from β-amyrin, oleanolic acid and the intermediates erythrodiol and oleanolic aldehyde. UGTs specific for the glucosylation of triterpenoids at the *C*-3 hydroxyl and *C*-28 carboxyl groups are expressed in strains engineered to produce β-amyrin or oleanolic acid. In addition, the co-expression of two UGTs was explored for the production of oleanane

saponin diglycosides. To date, only monoglycosides of oleanane-type triterpenoids have been produced in yeast (Moses et al., 2014b; Arendt et al., 2017).

Chapter 2: Materials and Methods

2.1 Materials

2.1.1 Yeast growth

All OD_{600} measurements were performed using a WPA biowave CO8000 cell density meter, unless otherwise stated.

All media were autoclaved prior to use.

YPD: 10 g/L yeast extract, 20 g/L peptone, 20 g/L D-glucose (media kitchen)

SC-URA + 2 % D-Glc: 6.9 g/L yeast nitrogen base (Sigma Y0626), 0.77 g/L CSM-URA (Formedium DCS0169), 20 g/L D-glucose

SC-URA + 2 % D-Gal: 6.9 g/L yeast nitrogen base (Sigma Y0626), 0.77 g/L CSM-URA (Formedium DCS0169), 20 g/L D-galactose

SC-URA-LEU + 2 % D-Glc: 6.9 g/L yeast nitrogen base (Sigma Y0626), 0.67 g/L CSM-LEU-URA (Formedium DCS0581), 20 g/L D-glucose

SC-URA-LEU + 2 % D-Gal: 6.9 g/L yeast nitrogen base (Sigma Y0626), 0.67 g/L CSM-LEU-URA (Formedium DCS0581), 20 g/L D-galactose

YPD agar: 10 g/L yeast extract, 20 g/L peptone, 20 g/L D-glucose (Fisher Scientific, G/0500/60), 20 g/L agar (Oxoid LP0011)

SC-URA + 2 % D-Glc agar: 6.9 g/L yeast nitrogen base (Sigma Y0626), 0.77 g/L CSM-URA (Formedium DCS0169), 20 g/L D-glucose (Fisher Scientific, G/0500/60), 20 g/L agar (Oxoid LP0011)

SC-URA-LEU + 2 % D-Glc agar: 6.9 g/L yeast nitrogen base (Sigma Y0626), 0.67 g/L CSM-LEU-URA (Formedium DCS0581), 20 g/L D-glucose (Fisher Scientific, G/0500/60), 20 g/L agar (Oxoid LP0011)

2.1.2 *E. coli* growth

All OD_{600} measurements were performed using a WPA biowave CO8000 cell density meter, unless otherwise stated

All media was autoclaved prior to use.

LB: 10 g/L tryptone, 5 g/L yeast extract, 10 g/L NaCl

LB agar: 10 g/L tryptone, 5 g/L yeast extract, 10 g/L NaCl. 20 g/L agar

2.1.3 Antibiotics

Ampicillin: stock concentration was 100 mg/ml in dH_2O; working concentration was 100 µg/ml

Chloramphenicol: stock concentration was 40 mg/ml in 100 % ethanol; working concentration was 40 µg/ml

Spectinomycin: stock concentration was 100 mg/ml in dH_2O; working concentration was 100 µg/ml

2.1.4 Molecular biology

TFB1 buffer: 30 mM CH_3COOK, 100 mM RbCl, 10 mM $CaCl_2$, 50 mM $MnCl_2$, 15% (v/v) glycerol; pH adjusted to 5.8 with acetate; sterilised by filtration (0.22 µm)

TFB2 buffer: 10 mM MOPS (pH 6.5), 75 mM $CaCl_2$, 10 mM RbCl, 15% (v/v) glycerol; pH adjusted to 6.5 with KOH; sterilised by filtration (0.22 µm)

QIAprep Spin Miniprep Kit: purchased from QIAGEN

E.Z.N.A.® Plasmid Mini Kit I, (Q-spin): purchased from Omega Bio-Tek

OneTaq: OneTaq® Quick-Load® 2X Master Mix with Standard Buffer was purchased from NEB

GoTaq: 2X GoTaq Green Master Mix was purchased from Promega

Phusion: 5X High-Fidelity Buffer, 2 mM dNTPs and 2 U/µl Phusion DNA Polymerase were purchased from NEB

Q5: 5X Q5 Reaction Buffer, 2 mM dNTPs and 2 U/µl Q5 High-Fidelity DNA Polymerase were purchased from NEB

Restriction enzymes: all were purchased from NEB

Agarose: purchased from Invitrogen

TAE (1X): 40 mM Tris, 20 mM acetate, 1 mM EDTA

QIAquick Gel Extraction Kit: purchased from QIAGEN

10X T4 DNA ligase buffer: purchased from NEB (B0202S)

T7 DNA ligase: purchased from NEB (M0318)

2.1.5 Yeast transformation

Salmon sperm DNA (10 mg/ml): purchased from Invitrogen (15632-011)

Lithium acetate (1 M): purchased from Alfa Aesar

PEG 3350, 50 % (w/v): 50 g PEG 3350 (Sigma 88276) dissolved in 100 ml dH_2O.

2.1.6 Yeast DNA extraction

Chelex 100: purchased from Sigma

Glass beads: purchased from Sigma

2.1.7 Chromatography standards

Coprostanol-3-ol: 50 mg coprostan-3-ol (≥ 98 %, Sigma C7578) dissolved in 100 % ethanol to a concentration of 2.57 mM (stored at -20 °C)

Ergosterol: 100 mg ergosterol (≥ 95 %, Sigma 45480-10G-F) dissolved in 100 % methanol to a concentration of 1 g/L (stored at -20 °C)

β-amyrin: 10 mg β-amyrin (Extrasynthese 0016 S) dissolved in 100 % methanol to a concentration of 1 g/L (stored at -20 °C)

Erythrodiol: 10 mg erythrodiol (Extrasynthese 0040 S) dissolved in 100 % methanol to a concentration of 2.5 g/L (stored at -20 °C)

Oleanolic acid: 20 mg oleanolic acid (Extrasynthese 0041 S) dissolved in 100 % methanol to a concentration of 0.5 g/L (stored at -20 °C)

3-*O*-Glc-echinocystic acid: 10 mg 3-*O*-Glc-echinocystic acid (Extrasynthese 0110 S) dissolved in 100 % methanol to a concentration of 100 mg/L (stored at -20 °C)

2.1.8 Metabolite extractions

Lysis solution: 20 % KOH, 50 % ethanol

Lysis solution with coprostanol: 20 % KOH, 50 % ethanol, 80 nmol/ml coprostan-3-ol

***N*-Hexane (≥ 99 %):** Honeywell Riedel-de Haën 32293

Ethyl acetate (anhydrous, 99.8 %): Sigma 270989

Butan-1-ol (≥ 99 %): Acros Organics 167695000

2.1.9 Thin-layer chromatography (TLC)

TLC plates: Silica gel on TLC Alu foils with fluorescent indicator 254 nm, silica gel matrix, L × W 5 cm × 10 cm (Sigma 91835-50EA)

TLC solvent (triterpenoid extracts): hexane : ethyl acetate (9:1)

TLC solvent (saponin extracts): chloroform : methanol : H_2O (30:6:1)

Stain (triterpenoid and saponin extracts): glacial acetic acid : p-anisaldehyde : concentrated sulphuric acid (48:1:1)

2.1.10 GC-MS and LC-MS

1.1 ml autosampler vials: 1.1 ml microliter crimp neck vial ND11, 32 x 11.6 mm, clear glass, 1st hydrolytic class; Gerstel 093640-121-00

Crimp caps: 11mm autosampler vial crimp caps with rubber or silicone liner (Thermo Scientific 11-MC-8RT1 or Thermo Scientific 11-MC-ST15)

MSTFA (*N*-Methyl-*N*-(trimethylsilyl)trifluoroacetamide with 1% trimethylchlorosilane): Sigma 69478

Pyridine (anhydrous, 99.8 %): Sigma 270970

LC-MS buffers: formic acid and methanol purchased from VWR

2.1.11 Proteomics

Trypsin: purchased from Worthington Labs

Omix C18 stage tips: purchased from Agilent

Nano-HPLC (Dionex Ultimate 3000 RSLC, Thermo-Fisher) coupled to a QExactive mass spectrometer (Thermo-Fisher) with a 300 µm x 5 mm pre-column (Acclaim Pepmap, 5 µm particle size) joined with a 75 µm x 50 cm column (Acclaim Pepmap, 3 µm particle size).

2.1.12 Microplate assays

96-well plates: Costar 3628

Tape: electrical insulation tape, 12mm x 20m, RS Components 504-2380

Tecan Sunrise plate reader

2.1.13 SDS-PAGE

Resolving gel (10 % polyacrylamide): 10 % bisacrylamide, 1X resolving gel buffer, 0.1 % (w/v) SDS, 0.0005 % (v/v) TEMED, 0.075 % (v/v) ammonium persulphate

Stacking gel (5.7 % polyacrylamide): 5.7 % bisacrylamide, 1X stacking gel buffer, 0.1 % (w/v) SDS, 0.001 % (v/v) TEMED, 0.05 % (v/v) ammonium persulphate

4X resolving gel buffer: 1.5 M Tris base, pH 8.8

4X stacking gel buffer: 0.5 M Tris base, pH 6.8

1X running buffer: 3 g/L Tris base, 14.4 g/L glycine, 0.1 % (w/v) SDS.

Lyse and load buffer: 50 mM Tris-HCl (pH 6.8), 4 % SDS, 8 M urea, 30 % glycerol, 0.1 M DTT, 0.005 % bromophenol blue

2.1.14 Western blots

Transfer buffer: 3.03 g/L Tris base, 14.4 g/L glycine, 20 % methanol

PBS-T: 11.5 g/L Na_2HPO_4, 2.96 g/L NaH_2PO_4, 5.84 g/L NaCl, 0.1 % Tween20, pH 7.5

Blocking solution: 3 % milk powder in PBS-T

Membrane: PVDF membrane purchased from Bio-Rad

Transfer machine: BioRad Mini Trans-Blot® Electrophoretic Transfer Cell

Primary antibody: mouse anti-His (Sigma H1029)

Secondary antibody: peroxidase-conjugated rabbit anti-mouse (Sigma A9044)

WesternSure Pen (LI-COR)

WesternSure Premium Chemiluminescent Substrate (LI-COR 926-95000)

LI-COR C-DiGit Blot Scanner

2.1.15 Reverse transcription-polymerase chain reaction (RT-PCR)

QIAGEN RNeasy Mini kit: purchased from QIAGEN

DNase I: 10X DNase I reaction buffer, DNase I and 25 mM EDTA (pH 8) purchased from Invitrogen (18068-015)

SuperScript III Reverse Transcriptase Kit: purchased from Invitrogen (18080)

Oligo(dT)$_{20}$ primer: purchased from Invitrogen (18418020)

dNTPs: purchased from NEB

SUPERase-In RNase Inhibitor (20 U/µL): purchased from Invitrogen (AM2694)

***E. coli* RNase H:** purchased from Invitrogen (18021071)

2.1.16 BioLector fermentations

FlowerPlate with pH and DO optodes: purchased from m2p-labs (MTP-48-BOH)

Gas-permeable sealing foil: purchased from m2p-labs

RoboLector XL (m2p-labs)

2.2 Methods

2.2.1 Storage of *E. coli* and yeast

For long term storage, *E. coli* and yeast strains were stored at -80 °C in 20 % (v/v) glycerol. For short term storage, strains were kept on agar plates at 4 °C.

To prepare *E. coli* glycerol stocks, 5 ml LB containing the appropriate antibiotic was inoculated with the desired *E. coli* strain and grown at 37 °C for ~ 16 h (shaking at 200 rpm). To prepare yeast glycerol stocks, 5 ml yeast growth medium was inoculated with the desired yeast strain and grown at 30 °C for ~ 24 h (shaking at 200 rpm). Yeast strains carrying no plasmids were grown in YPD, while strains with plasmids were grown in either SC-URA + 2 % D-Glc or SC-URA-LEU + 2 % D-Glc as appropriate. 500 µl of culture was mixed with 500 µl 40 % (v/v) glycerol and stored at -80°C.

2.2.2 Preparation of chemically competent cells

Chemically competent cells were prepared using the RbCl method. 5 ml of LB was inoculated with *E. coli* TOP10 and grown at 37 °C for ~ 16 h (shaking at 200 rpm). The culture was diluted 1:400 in 500 ml LB containing 20 mM $MgSO_4$ and grown at 37 °C (shaking at 200 rpm) until the OD_{600} reached ~ 0.5. The cells were transferred to ten 50 ml Falcon tubes (50 ml per tube) and incubated on ice for 10 min. The cells were kept cold from this point on. The cells were centrifuged (3750 RCF for 5 min at 4 °C), each was resuspended in 20 ml cold TFB1 buffer, incubated on ice for 5 min, centrifuged again (1350 RCF for 10 min at 4 °C), resuspended in 2 ml cold TFB2 buffer, and incubated on ice for 15 min. Cells were aliquoted, snap-frozen on dry ice and 100 % ethanol, and stored at -80°C.

2.2.3 Plasmid preparation

5 ml of LB containing the appropriate antibiotic was inoculated with a colony of *E. coli* containing the desired plasmid, and grown at 37 °C for ~ 16 h (shaking at 200 rpm). Plasmid extraction was performed using the QIAprep Spin Miniprep Kit or E.Z.N.A. Plasmid Mini Kit I according to manufacturer's instructions, eluting in 30 µl dH_2O.

2.2.4 Polymerase chain reaction (PCR)

PCR reactions to construct parts for cloning into pYTK001 were performed using Phusion DNA Polymerase (constructing the *GAL* promoters) or Q5 DNA Polymerase (removing the transposon from the *CqBAS1* gene; constructing the catalytically

inactive AaBAS). For Phusion PCR, the reaction was 1X HF buffer, 0.2 mM dNTP (NEB), 0.5 µM each primer, 0.2 µl template and 1 U Phusion DNA Polymerase. The thermocycling programme was: 98 °C for 30 s; 35 cycles of 98 °C for 10 s, anneal for 30 s, extension at 72 °C; 72 °C for 10 min; hold at 16 °C. For Q5 PCR, the reaction was 1X Q5 reaction buffer, 0.2 mM dNTPs, 0.5 µM each primer, 0.2 µl template and 1 U Q5 High-Fidelity DNA Polymerase. The thermocycling programme was: 98 °C for 30 s; 30 cycles of 98 °C for 10 s, anneal for 30 s, extension at 72 °C; 72 °C for 2 min; hold at 16 °C. In both cases the annealing temperature was 3 °C above the melting temperature (T_m) of the lower T_m primer, and the extension time was 30 s / kb PCR product. The PCR products were purified by gel extraction.

2.2.5 Restriction digest

Restriction digests of plasmid minipreps were performed as follows. Mixed 0.5-1.5 µl plasmid, 0.5 µl each restriction enzyme (NEB) and 1X CutSmart buffer (NEB) in a final volume of 20 µl. Incubated at 37 °C for at least 1 h and analysed by agarose gel electrophoresis.

2.2.6 Agarose gel electrophoresis

Restriction digests, PCR reactions or total RNA were mixed with 1X Gel Loading Dye, Purple (NEB B7024S) and loaded onto an agarose-TAE gel. PCR reactions that used GoTaq or OneTaq were loaded directly onto the gel. The percentage of agarose was typically 1 % but was adjusted based on the expected size of the bands. The gel was run at 100 V in 1X TAE buffer and imaged using a Gel Doc™ XR+ Gel Documentation System (BioRad).

2.2.7 Gel extraction

PCR reactions were run on agarose-TAE gels and purified using the QIAquick Gel Extraction Kit, according to the manufacturer's instructions. Elution was in 30 µl dH_2O.

2.2.8 Sanger sequencing

Plasmids were sequenced by MRC PPU DNA Sequencing and Services (University of Dundee) using the Sanger sequencing method.

2.2.9 Plasmid construction

Plasmids were assembled using MoClo-YTK (Lee et al., 2015). Briefly, individual parts (e.g. promoters and open reading frames) are first cloned into an entry vector

(pYTK001), transcriptional units (promoter-ORF-terminator) are then constructed by cloning these parts into a pCP series plasmid, and pathways are finally constructed by cloning multiple transcriptional units into a pMG series plasmid.

All genes except AsBAS were codon-optimised for *S. cerevisiae*, synthesised (GeneArt, Gen9, or IDT) and cloned into the MoClo-YTK entry vector (pYTK001). For AsBAS, a wild-type sequence was provided by Thomas Louveau (Anne Osbourn lab, John Innes Centre); PCR was performed to add appropriate BsmBI restriction sites and the product was cloned into pYTK001. To make dAaBAS, PCR was performed on AaBAS (in pYTK001) to introduce a D484N mutation (G1450A nucleotide) and the product was cloned into pYTK001. To generate a transposon-free CqBAS1, PCR was performed on CqBAS1 synthetic DNA to introduce silent mutations (C552T, G555A, C570A nucleotides) and the product was cloned into pYTK001. The GAL1 and GAL10 promoters were made by performing PCR on the GAL1/GAL10 bidirectional promoter of pESC-URA (using primer pairs PGAL1 fwd/PGAL1 rev and PGAL10 fwd 1/PGAL10 rev) and cloning the product into pYTK001. The GAL7 and GAL3 promoters were made by performing PCR on *S. cerevisiae* BY4741 genomic DNA (using primer pairs (MoClo)-PGAL3 fwd/(MoClo)-PGAL3 rev and (MoClo)-PGAL7 fwd/(MoClo)-PGAL7 rev), respectively and cloning the product into pYTK001. All other parts were provided with the MoClo-YTK.

For the BAS comparisons (Chapter 4), each BAS was cloned into the plasmid pCPS1ULA (CEN6/ARS4, URA3) with the GAL1 promoter and TDH1 terminator, generating the pCPS1ULA-BAX series. For the CYP716A comparisons (Chapter 5), each CYP716A was cloned into pCP12ULA (CEN6/ARS4, URA3) with the GAL10 promoter and the ENO2 terminator, and ATR2 was cloned into pCP2EULA (CEN6/ARS4, URA3) with the GAL3 promoter and the PGK1 terminator. Plasmids encoding the appropriate BAS (pCPS1ULA-BAS), CYP716A (pCP12ULA-CYP716A) and ATR2 (pCP2EULA-ATR2) were used to clone these genes into pMGSEULS (CEN6/ARS4, URA3), generating the pMGSEULS-OAX series of plasmids. For saponin production (Chapter 6), UGTs were cloned into a low copy (CEN6/ARS4) pCP plasmid with LEU2 selection marker under the control of either the GAL1, GAL10 or GAL7 promoter. For the co-expression of UGT73C10, and UGT72P2, both transcriptional units were cloned into a pMG plasmid (CEN6/ARS4, LEU2).

To clone parts into pYTK001, the appropriate PCR product or synthetic DNA was mixed with 20 fmol pYTK001, 1X T4 DNA ligase buffer (NEB), 1500 U T7 DNA ligase (NEB), and 5 U BsmBI (NEB) for a final volume of 10 µl. The reaction was thermocycled using the following programme: 42 °C for 2 min then 16 °C for 5 min (25 cycles); 60 °C for 10 min; 80 °C for 10 min; hold at 16 °C. Assemblies into pCP series vectors were performed similarly, except that the reaction contained 20 fmol each plasmid (the pCP plasmid and the parts in pYTK001), and 5 U BsaI (NEB) was used instead of BsmBI. Assemblies into pMG series vectors were performed the same as for pCP vectors, but the restriction enzyme was 5 U BsmBI.

Assemblies were transformed into *E. coli* TOP10. 5 µl of reaction was added to 50 - 200 µl of chemically competent cells, the cells were incubated on ice for 30 min, at 42 °C for 30 s, then on ice for 2 min. 250 - 500 µl SOC media was added and the cells were incubated at 37 °C for 90 min (shaking at 200 rpm). 100 µl was plated onto LB-agar containing the appropriate antibiotic (40 µg/ml chloramphenicol for assemblies into pYTK001, 100 µg/ml ampicillin for pCP plasmids, and 100 µg/ml spectinomycin for pMG plasmids), and incubated at 37 °C until colonies had appeared (usually ~ 16 h). Plasmids were prepared by miniprep and verified by restriction digest and Sanger sequencing.

2.2.10 Yeast transformation

pCP and pMG plasmids encoding appropriate transcription units / pathways were transformed into yeast to generate triterpenoid/saponin production strains. For β-amyrin and oleanolic acid production (Chapters 4-5), *S. cerevisiae* BY4741 was transformed with the appropriate pCP or pMG plasmids (CEN6/ARS4, URA3). For the production of saponins (Chapter 6), *S. cerevisiae* BY4741 was first transformed with pCPS1ULA-BA5 (CEN6/ARS4, URA3; encoding EtBAS) or pMGSEULS-OA7 (CEN6/ARS4, URA3; encoding EtBAS, CYP716A12 and ATR2) to generate strains MD-BA5 and MD-OA7, respectively. These strains were then transformed with pCP or pMG plasmids (CEN6/ARS4, LEU2) encoding the appropriate UGTs.

Yeast transformation was performed as follows. Cells were grown for 24 h in appropriate media (YPD for *S. cerevisiae* BY4741, or appropriate drop-out media for strains already carrying a plasmid), diluted into the same media (10 ml per transformation) for an OD_{600} of 0.1, and grown until OD_{600} 0.4-0.6 (~ 5 h). The culture was transferred to 50 ml Falcon tubes, pelleted (3000 RCF for 5 min), resuspended

in half volume dH_2O, pelleted again (3000 RCF for 2 min) and resuspended in 1/10th volume dH_2O. Cells were transferred to 1.5 ml microtubes (1 ml / tube), spun at 8000 rpm for 1 min, and resuspended in 54 ul dH_2O. Meanwhile, salmon sperm DNA was incubated at 100 °C for 8 min. 100-500 ng plasmid and 306 µl of transformation mix (10 µl 10 mg/ml salmon sperm DNA, 36 µl 1 M lithium acetate and 260 µl 50 % PEG 3350) was added, and cells were mixed by vortexing briefly. Cells were incubated at 30 °C for 30 min, 42 °C for 14 min, pelleted (8000 rpm, 2 min) and resuspended in 100 µl dH_2O, and 100 µl was plated onto appropriate selective media and incubated at 30 °C for 48 h.

2.2.11 Yeast strain verification

Individual colonies from the transformation were re-streaked and incubated at 30 °C for 48 h. DNA was extracted using the GC prep method described by Blount *et al.* (Blount et al., 2016). Cells were picked from the plate, resuspended in 100 µl 5 % Chelex 100, and glass beads were added to approximately half sample volume. Cells were vortexed at maximum speed for 4 min, incubated at 100 °C for 2 min, pelleted (14,000 rpm, 1 min), and the supernatant (containing the DNA) was transferred to a fresh tube.

The extracts were analysed by PCR using a touch-down protocol. The reaction was set up as follows: 0.4 µl 5 µM each primer, 0.2 µl DNA extract, 5 µl OneTaq Quick-Load 2X Master Mix with Green Buffer (NEB M0485), 5 µl dH_2O. The cycling programme was: 94 °C for 30 s; 20 cycles of 94 °C for 30 s, 70 °C for 45 s (decreasing by 1 C per cycle), 68 °C for 1 min; 15 cycles of 94 °C for 30 s, 55 °C for 45 s, 68 °C for 1 min; 68 °C for 5 min; hold at 16 °C. 5 ul of PCR reaction was analysed by agarose gel electrophoresis.

2.2.12 Triterpenoid extractions for GC-MS analysis

Precultures were grown in 50 ml Falcon tubes by inoculating 5 ml SC-URA + 2 % D-Glc with the desired yeast strain and incubating at 30 °C for 24 h (shaking at 200 rpm). The precultures were washed twice in 5 ml dH_2O (for each wash spun at 3000 RCF for 2 min, then resuspended in 5 ml dH_2O), diluted to an OD_{600} of 0.2 in 5 ml SC-URA + 2 % D-Gal in 55 ml glass culture vials, and incubated at 30 °C for 96 h (shaking at 200 rpm). The OD_{600} was measured, then 1 ml of yeast culture was transferred to a microtube and pelleted (14,000 rpm, 1 min), and washed twice in 1 ml dH_2O (spun at

14,000 rpm for 1 min, resuspended in 1 ml dH_2O, then spun again). The pellet was resuspended in 0.5 ml lysis solution containing 80 nmol/ml coprostan-3-ol and incubated at 100 °C for 10 min (with the lids open to allow the ethanol, which is miscible with ethyl acetate, to evaporate). The cells were then extracted twice with 500 µl hexane then twice with ethyl acetate. Extractions were performed by adding 500 µl hexane or ethyl acetate, vortexing briefly at max speed, centrifuging 14,000 rpm for 1 min, then removing the organic phase to a separate microtube. For the first extraction 400 µl organic phase was removed and for the subsequent extractions 500 µl was removed, and these were pooled together to give a total of 1900 µl extract. 600 µl of extract was transferred to a 1.1 ml autosampler vial, evaporated to dryness using a GeneVac EZ-2 Elite evaporator (Low BP Mix programme, 40 °C), and sealed with crimp caps. The dried extracts were stored at room temperature until analysis.

2.2.13 Triterpenoid extractions for TLC analysis

Cultures were grown as for the triterpenoid extractions for GC-MS analysis, except that growth in SC-URA + 2 % D-Gal was done in 50 ml Falcon tubes for 90 h. The OD_{600} was measured, and 1 ml of culture was pelleted (14,000 rpm, 1 min) and the supernatant was removed. Cells were resuspended in 0.5 ml lysis solution and incubated at 75 °C for 2 h. The extract (supernatant and cell debris) was transferred a 7 ml glass vial, mixed with 2 ml dH_2O, and extracted twice with 1 ml hexane. Extractions were performed by adding 1 ml hexane, vortexing at maximum speed briefly, letting the aqueous and organic phases separate, transferring the aqueous phase to a fresh glass vial, and extracting the aqueous phase again. The organic phases and any emulsions were pooled together, washed with 1 ml 20 % NaCl to separate the emulsions (added 1 ml 20 % NaCl, vortexed briefly at maximum speed, let the phases separate and removed the aqueous phase). The pooled organic phases were evaporated to dryness overnight (at room temperature and atmospheric pressure).

2.2.14 Saponin extractions for LC-MS analysis

Cells were grown as for the triterpenoid extractions for GC-MS analysis, except that growth in SC-URA + 2 % D-Gal was for 90 h. The OD_{600} was measured and 1 ml of culture was transferred to a microtube and pelleted (14,000 rpm, 1 min). 500 µl supernatant (the “spent medium”) was transferred to a fresh microtube and the pellet was washed in 1 ml dH_2O (resuspended pellet in 1 ml dH_2O then spun at 14,000 rpm

for 1 min). The pellet was resuspended in 100 µl TN150-Triton lysis buffer, ~ 250 µl glass beads were added, and the cells were lysed by vortexing (three cycles of vortexing at maximum speed for 1 min followed by incubating on ice for 30 s). 900 µl dH_2O was added, then spun at 14,000 rpm for 1 min and transferred 500 µl supernatant to a fresh microtube. The lysed cells and spent medium were extracted thrice with butan-1-ol. For each extraction, added 500 µl butan-1-ol, vortexed briefly at maximum speed, spun at 14,000 rpm for 1 min, and transferred the butan-1-ol phase to a separate microtube (the butan-1-ol phases for the three extractions were pooled together). 450 µl of extract was transferred to a 1.1 ml autosampler vial, evaporated to dryness in a GeneVac EZ-2 Elite evaporator (High BP programme, 50 °C), and sealed with crimp caps. The dried extracts were stored at -20 °C until analysis.

2.2.15 Saponin extractions for TLC analysis

Cells were grown as for the triterpenoid extractions for TLC analysis. The OD_{600} was measured and 1 ml of culture was transferred to a microtube and pelleted (14,000 rpm, 1 min). 700 µl supernatant (the “spent medium”) was transferred to a fresh microtube and the pellet was washed twice in 1 ml dH_2O (a single wash consisted of resuspending the pellet in 1 ml dH_2O then centrifuging at 14,000 rpm for 1 min). The pellet was resuspended in 100 µl TN150-Triton lysis buffer, glass beads were added to ~ 250 µl volume, and the cells were lysed by vortexing (three cycles of vortexing at maximum speed for 1 min followed by incubating on ice for 30 s). 1 ml dH_2O was added, the cells were vortexed briefly at maximum speed, centrifuged (14,000 rpm, 1 min), and 700 µl supernatant was transferred to a 2 ml microtube. The lysed cells and spent medium was extracted twice with 700 µl butan-1-ol, and the pooled extracts were evaporated to dryness using a SpinVac.

2.2.16 Thin Layer Chromatography (TLC)

Dried extracts were dissolved in 100 µl hexane (triterpenoid extracts) or 100 µl butan-1-ol (saponin extracts) by vortexing at maximum speed. 10 µl of each extract and 0.6 µg of authentic standards were loaded onto a TLC plate. This was done by adding 2 µl of extract or 100 mg/L standard at a time, allowing the plate to dry between each addition, until 10 µl or 0.6 µg had been loaded. The plate was placed in a TLC tank containing hexane:ethyl acetate (9:1) for the triterpenoid extracts, or containing chloroform : methanol : H_2O (30:6:1) for the saponin extracts. The tank was sealed

and the plate was removed once the solvent front had reached the top of the plate. The plate was allowed to dry, stained in glacial acetic acid : p-anisaldehyde : concentrated sulphuric acid (48:1:1), then incubated at 100 °C until purple spots appeared.

2.2.17 Gas chromatography-mass spectrometry (GC-MS)

Triterpenoid extracts were analysed using an Agilent 7200B GC/Q-TOF with Gerstel Multi-Purpose Sampler (MPS). Samples were derivatised by trimethylsilylation: 50 µl of MSTFA:pyridine (4:1) was added to the dried extract, vortexed for 30 s at 3000 rpm, incubated at 37 °C for 30 min (shaking at 750 rpm), and centrifuged at 4500 rpm for 2 min. Derivatisation was performed by the MPS and samples were derivatised sequentially, with each sample being injected immediately after derivatisation. Injection was at 1 µl volume with a 10:1 split. Where analyte peak areas were outside the linear range of the standard curve, the sample was injected again at a 5:1 split or a 25:1 split. The column was an Agilent DB-5ms column (30 m with 10 m guard x 250 µm x 0.25 µm). The GC parameters were as follows: constant helium flow of 1 ml / min; the inlet was set to 250 °C; the oven was initially held at 100 °C for 2 min, then increased to 300 °C at a rate of 25 °C / min, then held at 300 °C for 25 min. Ionisation was by electron ionisation (EI), and the MS parameters were: 230 °C source temperature, 35 µA filament current, 70 eV electron energy, mass range of 60-900 m/z, acquisition rate of 4 spectra / s, 13 min solvent delay. Total ion chromatograms and mass spectra were analysed using the Agilent MassHunter Qualitative Analysis B.07.00 software, and peak areas were calculated using the Agile 2 integrator method. Hannah Florance (University of Edinburgh) helped develop the protocol and provided training on using the instrument.

2.2.18 Liquid chromatography-mass spectrometry (LC-MS)

Saponin extracts were analysed using an Agilent 6560 IM Q-TOF LC/MS coupled to an Agilent 1290 Infinity II UHPLC. Dried extracts were reconstituted in 100 µl methanol (vortexed briefly at maximum speed then centrifuged briefly at 500 RCF). Samples were injected at 10 µl volume onto an ACE C18 PFP column (100mm x 2.1mm, 1.8 µm). The gradient was as follows: 70 % buffer A and 30 % buffer B at 0-1 min, 100 % buffer B at 10-12 min, 70 % buffer A and 30 % buffer B at 12.1 min; buffer A was H_2O with 0.1 % formic acid, and buffer B was methanol with 0.1 % formic acid. The column was held at 30 °C and a constant flow rate of 0.2 ml / min was used. For the mass

spectrometry, ionisation was by electrospray ionisation (ESI), the acquisition mode was ion mobility-QTOF, and positive ion mode was used. The MS parameters were as follows: the gas temperature was 325 °C, the drying gas was set to 13 L / min, the nebuliser was set to 20 psig, the sheath gas temperature and flow were 275 °C and 12 L / min, the VCap was 3500 V, the nozzle voltage was 700 V, the fragmenter was 400 V, the collision energy was 65 V, the mass range was 100-1600 m/z, and the acquisition rate was 1 frame / s. Chromatograms and mass spectra were analysed using the Agilent MassHunter Qualitative Analysis B.07.00 software. Hannah Florance (University of Edinburgh) helped develop the protocol and provided training on using the instrument.

2.2.19 Microplate assays

The washed precultures used for the triterpenoid extractions were diluted to an OD_{600} of 0.2 in 100 µl SC-URA + 2 % D-Glc or 100 µl SC-URA + 2 % D-Gal in 96-well plates. Blanks contained only 100 µl SC-URA + 2 % D-Glc. A lid was placed on the plate and sealed with tape, and OD_{595} was read on a Tecan Sunrise plate reader every 15 min for 96 h. Temperature was 30 °C, shaking mode was “inside” and shaking intensity was “normal”. Data was exported into Microsoft Excel and analysed in MATLAB.

2.2.20 Western blots

Precultures were grown as for the triterpenoid extractions, diluted into 5 ml SC-URA + 2 % D-Glc or 5 ml SC-URA + 2 % D-Gal for an OD_{600} of 0.2, and grown at 30 °C for 48 h (shaking at 200 rpm). 1 ml of culture was pelleted (14,000 rpm, 1 min) and resuspended in an appropriate volume of lyse and load buffer such that the OD_{600} was 100. The cells were lysed by incubating at 100 °C for 5 min, then 10 µl was immediately loaded onto a 10 % polyacrylamide gel. 8 µl of colour prestained protein standard broad range (NEB P7712) was also loaded. The gel as run at 80 V until the samples were through the stacking gel, then at 150 V for ~ 45 min (until the 30 kDa marker had reached the bottom of the resolving gel).

The gel was blotted onto a PVDF membrane using a BioRad Mini Trans-Blot® Electrophoretic Transfer Cell (100 V for 60 min in cold transfer buffer), and the membrane was blocked overnight in blocking solution. The primary antibody was mouse anti-His (Sigma H1029) and the secondary antibody was peroxidase-conjugated rabbit anti-mouse (Sigma A9044). Antibody incubations were performed

with 1:1500 concentration of primary antibody or a 1:15,000 concentration of secondary antibody in blocking solution, incubating at room temperature for 1 h (shaking gently). After each incubation the membrane was washed three times with excess PBS-T. The ladder was marked on the membrane using a WesternSure Pen (LI-COR), and the membrane was imaged by chemiluminescence using the WesternSure Premium Chemiluminescent Substrate (LI-COR 926-95000) on a LI-COR C-DiGit Blot Scanner.

2.2.21 Reverse transcription-polymerase chain reaction (RT-PCR)

Precultures were grown as for the triterpenoid extractions, diluted to an OD_{600} of 0.1 in 5 ml SC-URA + 2 % D-Gal, and grown at 30 °C (shaking at 200 rpm) until they reached an OD_{600} of ~ 0.5. Total RNA was prepared using the QIAgen RNeasy Mini Kit, harvesting 1 ml of culture. For the strains expressing CYP716A52v2 and ATR2, 0.5 ml culture was harvested because the OD_{600} was ~ 1, while for the strains expressing PgBAS1 and GgBAS, 1.5 ml culture was harvested because the OD_{600} was ~ 0.3. Lysis was performed on freshly harvested cells using the enzymatic method for ≤ 2 x 10^7 cells. 500 ng of total RNA was analysed by agarose gel electrophoresis (using 1 % agarose-TAE gels) to confirm RNA integrity.

The RNA was treated with DNase I (Invitrogen 18068-015): mixed 200 ng RNA, 1X DNase I reaction buffer, and 1 µl DNase I in 10 µl final volume; incubated 15 min at room temperature; added 1 µl 25 mM EDTA (pH 8); incubated 10 min at 65 °C. Reverse transcription was then performed using the SuperScript III Reverse Transcriptase Kit (Invitrogen 18080): mixed 1 µl of 50 µM oligo$(dT)_{20}$, 1 µl of 10 mM dNTPs, and 10 µl of DNase I reaction; incubated at 65 °C for 5 min then on ice for 1 min; added 4 µl 5X First-Strand buffer, 1 µl 0.1 M DTT, 2 µl 20 U/µl SUPERase-In RNase Inhibitor, and 1 µl 200 U/µl SuperScript III RT (negative controls received 1 µl RNase-free water instead); incubated at 50 °C for 60 min, then at 70 °C for 15 min. 1 µl of *E. coli* RNase H (Invitrogen 18021071) was added to each reaction, and incubated at 37 °C for 20 min. PCR of the cDNA was then performed. Reactions contained 1X GoTaq Green Master Mix (Promega), 0.5 µM each primer, and 0.4 µl cDNA in 10 µl total volume. The thermocycling programme was: 95 °C for 5 min; 30 cycles of 95 °C for 45 s, 58 °C for 45 s, and 72 °C for 1 min; 72 °C for 5 min; hold at

16 °C. The PCR reactions were analysed by agarose gel electrophoresis using 1.5 % agarose-TAE gels.

2.2.22 BioLector fermentations

Precultures were grown by inoculating 5 ml SC-URA + 2 % D-Glc with the desired yeast strains and incubating at 30 °C for 24 h (shaking at 200 rpm). The cultures were washed twice in 5 ml dH_2O and diluted to an OD_{600} of 0.2 in 1.2 ml SC-URA + 2 % D-Gal in a 48-well FlowerPlate with pH and DO optodes (MTP-48-BOH). Blanks contained only 1.2 ml SC-URA + 2 % D-Gal. The plates were sealed with gas-permeable sealing foil, and cells were cultured in an m2p-labs RoboLector XL at 30 °C for 96 h (shaking at 1200 rpm). Humidity was maintained at 85 % relative humidity (rH), and biomass (with a gain of 8), pH and dissolved oxygen (DO) were measured every 15 min. At the end of the run, 900 µl of cells were harvested, OD_{600} was measured, and metabolites were extracted as described for the triterpenoid extractions for GC-MS analysis.

2.2.23 Phylogenetic analysis

Protein sequences were aligned using ClustalW and the phylogenetic tree was generated in MEGA7 (Kumar et al., 2016) using the maximum likelihood method, bootstrapping with 1000 replicates. The Poisson substitution model was used, rates among sites were assumed to be uniform, and gaps and missing data were removed using the complete deletion option. The maximum likelihood heuristic method was Nearest-Neighbour-Interchange and the initial tree was made using the neighbour joining method.

2.2.24 Proteomics

Cells were cultured as for the triterpenoid extractions, except that growth in SC-URA + 2 % D-Gal media was for 32 h. The OD_{600} was measured, 1 ml of culture was pelleted (14,000 rpm, 1 min), the supernatant was removed and pellets were stored at 4 °C until analysis. The pellets were resuspended in 8 M Urea and lysed using a Precellys Tissue Lyser. Protein concentration was determined by Bradford assay (Bio-Rad) and 50 µg was digested in 2 M Urea (with 25 mM Ammonium bicarbonate, 5 mM DTT and 12.5 mM Iodoacetamide) with 1 µg trypsin (Worthington Labs) at room temperature overnight. Samples were cleaned using Omix C18 stage tips (Agilent) and the eluted peptides were evaporated to dryness using a SpeedVac and stored at

-20 °C until analysis. Samples were dissolved in MS-loading buffer (0.05 % v/v trifluoroacetic acid in water) to 1 µg/µl final concentration and filtered using a Millex filter. The samples were mixed with an equal volume of 0.05 % TFA and 5 µl was injected for HPLC-MS analysis. Nano-ESI-HPLC-MS/MS analysis was performed using an online system of a nano-HPLC (Dionex Ultimate 3000 RSLC, Thermo-Fisher) coupled to a QExactive mass spectrometer (Thermo-Fisher) with a 300 µm x 5 mm pre-column (Acclaim Pepmap, 5 µm particle size) joined with a 75 µm x 50 cm column (Acclaim Pepmap, 3 µm particle size). A 90 min multi-step gradient of 2-98 % solvent B at a flow rate of 300 nL/min was used to separate the peptides; solvent A was 2 % acetonitrile in H_2O 0.1 % formic acid; solvent B was 80 % acetonitrile in H_2O with 0.1% formic acid. Cell lysis through to HPLC-MS analysis was performed by Lisa Imrie (Edinomics, University of Edinburgh).

2.3 List of yeast strains

Strain	Genotype	Notes
S. cerevisiae BY4741	*MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0*	
MD-BA1	BY4741, *[pCPS1ULA-BA1: CEN6/ARS4-ura3, pGAL1-AsBAS (JIC)-tTDH1]*	AsBAS
MD-BA2	BY4741, *[pCPS1ULA-BA2: CEN6/ARS4-ura3, pGAL1-PgBAS1 (synth)-tTDH1]*	PgBAS1
MD-BA3	BY4741, *[pCPS1ULA-BA3: CEN6/ARS4-ura3, pGAL1-GgBAS (synth)-tTDH1]*	GgBAS
MD-BA4	BY4741, *[pCPS1ULA-BA4: CEN6/ARS4-ura3, pGAL1-LjBAS (synth)-tTDH1]*	LjBAS
MD-BA5	BY4741, *[pCPS1ULA-BA5: CEN6/ARS4-ura3, pGAL1-EtBAS (synth)-tTDH1]*	EtBAS
MD-BA6	BY4741, *[pCPS1ULA-BA6: CEN6/ARS4-ura3, pGAL1-PtBAS (synth)-tTDH1]*	PtBAS
MD-BA9	BY4741, *[pCPS1ULA-BA9: CEN6/ARS4-ura3, pGAL1-GhBAS(Sc)-tTDH1]*	GhBAS
MD-BA14	BY4741, *[pCPS1ULA-BA14: CEN6/ARS4-ura3, pGAL1-AaBAS(Sc)-tTDH1]*	AaBAS
MD-BA15	BY4741, *[pCPS1ULA-BA15: CEN6/ARS4-ura3, pGAL1-BvBAS(Sc)-tTDH1]*	BvBAS
MD-BA16	BY4741, *[pCPS1ULA-BA16: CEN6/ARS4-ura3, pGAL1-CqBAS1(Sc)_2-tTDH1]*	CqBAS1
MD-BA17	BY4741, *[pCPS1ULA-BA17: CEN6/ARS4-ura3, pGAL1-MtBAS(Sc)-tTDH1]*	MtBAS
MD-BA18	BY4741, *[pCPS1ULA-BA18: CEN6/ARS4-ura3, pGAL1-SlBAS(Sc)-tTDH1]*	SlBAS
MD-OA1	BY4741, *[pMGSEULS-OA1: CEN6/ARS4-ura3, pGAL1-GgBAS(Sc)-tTDH1--pGAL10v1-CYP716A12(Sc)-tENO2--pGAL3-AtATR2(Sc)-tPGK1]*	CYP716A12 (+GgBAS)
MD-OA5	BY4741, *[pMGSEULS-OA5: CEN6/ARS4-ura3, pGAL1-GgBAS(Sc)-tTDH1--pGAL10v1-CYP716A147(Sc)-tENO2--pGAL3-AtATR2(Sc)-tPGK1]*	CYP716A147 (+GgBAS)
MD-OA6	BY4741, *[pMGSEULS-OA6: CEN6/ARS4-ura3, pGAL1-GgBAS(Sc)-tTDH1--pGAL10v1-CYP716A48(Sc)-tENO2--pGAL3-AtATR2(Sc)-tPGK1]*	CYP716A48 (+GgBAS)
MD-OA7	BY4741, *[pMGSEULS-OA7: CEN6/ARS4-ura3, pGAL1-EtBAS(Sc)-tTDH1--pGAL10v1-CYP716A12(Sc)-tENO2--pGAL3-AtATR2(Sc)-tPGK1]*	CYP716A12 (+EtBAS)
MD-OA13	BY4741, *[pMGSEULS-OA13: CEN6/ARS4-ura3, pGAL1-AaBAS(Sc)-tTDH1--pGAL10v1-*	CYP716A12 (+AaBAS)

	CYP716A12(Sc)-tENO2--pGAL3-AtATR2(Sc)-tPGK1]	
MD-OA14	BY4741, *[pMGSEULS-OA14: CEN6/ARS4-ura3, pGAL1-AaBAS(Sc)-tTDH1--pGAL10v1-CYP716A15(Sc)-tENO2--pGAL3-AtATR2(Sc)-tPGK1]*	CYP716A15 (+AaBAS)
MD-OA15	BY4741, *[pMGSEULS-OA15: CEN6/ARS4-ura3, pGAL1-AaBAS(Sc)-tTDH1--pGAL10v1-CYP716AL1(Sc)-tENO2--pGAL3-AtATR2(Sc)-tPGK1]*	CYP716AL1 (+AaBAS)
MD-OA16	BY4741, *[pMGSEULS-OA16: CEN6/ARS4-ura3, pGAL1-AaBAS(Sc)-tTDH1--pGAL10v1-CYP716A52v2(Sc)-tENO2--pGAL3-AtATR2(Sc)-tPGK1]*	CYP716A52v2 (+AaBAS)
MD-OA17	BY4741, *[pMGSEULS-OA17: CEN6/ARS4-ura3, pGAL1-AaBAS(Sc)-tTDH1--pGAL10v1-CYP716A147(Sc)-tENO2--pGAL3-AtATR2(Sc)-tPGK1]*	CYP716A147 (+AaBAS)
MD-OA18	BY4741, *[pMGSEULS-OA18: CEN6/ARS4-ura3, pGAL1-AaBAS(Sc)-tTDH1--pGAL10v1-CYP716A48(Sc)-tENO2--pGAL3-AtATR2(Sc)-tPGK1]*	CYP716A48 (+AaBAS)
MD-OA19	BY4741, *[pMGSEULS-OA19: CEN6/ARS4-ura3, pGAL1-AaBAS(Sc)-tTDH1--pGAL10v1-CYP716A44(Sc)-tENO2--pGAL3-AtATR2(Sc)-tPGK1]*	CYP716A44 (+AaBAS)
MD-OA20	BY4741, *[pMGSEULS-OA20: CEN6/ARS4-ura3, pGAL1-AaBAS(Sc)-tTDH1--pGAL10v1-CYP716A75(Sc)-tENO2--pGAL3-AtATR2(Sc)-tPGK1]*	CYP716A75 (+AaBAS)
MD-OA21	BY4741, *[pMGSEULS-OA21: CEN6/ARS4-ura3, pGAL1-AaBAS(Sc)-tTDH1--pGAL10v1-CYP716A79(Sc)-tENO2--pGAL3-AtATR2(Sc)-tPGK1]*	CYP716A79 (+AaBAS)
MD-OA22	BY4741, *[pMGSEULS-OA22: CEN6/ARS4-ura3, pGAL1-AaBAS(Sc)-tTDH1--pGAL10v1-CYP716A80(Sc)-tENO2--pGAL3-AtATR2(Sc)-tPGK1]*	CYP716A80 (+AaBAS)
MD-OA23	BY4741, *[pMGSEULS-OA23: CEN6/ARS4-ura3, pGAL1-AaBAS(Sc)-tTDH1--pGAL10v1-CYP716A83(Sc)-tENO2--pGAL3-AtATR2(Sc)-tPGK1]*	CYP716A83 (+AaBAS)
MD-OA24	BY4741, *[pMGSEULS-OA24: CEN6/ARS4-ura3, pGAL1-AaBAS(Sc)-tTDH1--pGAL10v1-CYP716A110(Sc)-tENO2--pGAL3-AtATR2(Sc)-tPGK1]*	CYP716A110 (+AaBAS)
MD-OA25	BY4741, *[pMGSEULS-OA25: CEN6/ARS4-ura3, pGAL1-AaBAS(Sc)-tTDH1--pGAL10v1-CYP716A140(Sc)-tENO2--pGAL3-AtATR2(Sc)-tPGK1]*	CYP716A140 (+AaBAS)

MD-OA26	BY4741, *[pMGSEULS-OA26: CEN6/ARS4-ura3, pGAL1-AaBAS(Sc)-tTDH1--pGAL10v1-CYP716A179(Sc)-tENO2--pGAL3-AtATR2(Sc)-tPGK1]*	CYP716A179 (+AaBAS)
MD-OA27	BY4741, *[pMGSEULS-OA27: CEN6/ARS4-ura3, pGAL1-AaBAS(Sc)-tTDH1--pGAL10v1-CYP716A244(Sc)-tENO2--pGAL3-AtATR2(Sc)-tPGK1]*	CYP716A244 (+AaBAS)
MD-OA28	BY4741, *[pMGSEULS-OA28: CEN6/ARS4-ura3, pGAL1-AaBAS(Sc)-tTDH1--pGAL10v1-CYP716A1(Sc)-tENO2--pGAL3-AtATR2(Sc)-tPGK1]*	CYP716A1 (+AaBAS)
MD-OA29	BY4741, *[pMGSEULS-OA29: CEN6/ARS4-ura3, pGAL1-AsBAS (JIC)-tTDH1--pGAL10v1-CYP716AL1(Sc)-tENO2--pGAL3-AtATR2(Sc)-tPGK1]*	CYP716AL1 (+AsBAS)
MD-OA30	BY4741, *[pMGSEULS-OA30: CEN6/ARS4-ura3, pGAL1-SlBAS(Sc)-tTDH1--pGAL10v1-CYP716AL1(Sc)-tENO2--pGAL3-AtATR2(Sc)-tPGK1]*	CYP716AL1 (+SlBAS)
MD-OA31	BY4741, *[pMGSEULS-OA31: CEN6/ARS4-ura3, pGAL1-SlBAS(Sc)-tTDH1--pGAL10v1-CYP716A52v2(Sc)-tENO2--pGAL3-AtATR2(Sc)-tPGK1]*	CYP716A52v2 (+SlBAS)
MD-BA1-t	BY4741, *[pCPS1ULA-BA1-t: CEN6/ARS4-ura3, pGAL1-AsBAS(JIC)v2-3xFLAG-6xHis-tTDH1]*	For western blot and RT-PCR of AsBAS
MD-BA2-t	BY4741, *[pCPS1ULA-BA2-t: CEN6/ARS4-ura3, pGAL1-PgBAS1(Sc)-3xFLAG-6xHis-tTDH1]*	For western blot and RT-PCR of PgBAS1
MD-BA3-t	BY4741, *[pCPS1ULA-BA3-t: CEN6/ARS4-ura3, pGAL1-GgBAS(Sc)-3xFLAG-6xHis-tTDH1]*	For western blot and RT-PCR of GgBAS
MD-BA4-t	BY4741, *[pCPS1ULA-BA4-t: CEN6/ARS4-ura3, pGAL1-LjBAS(Sc)-3xFLAG-6xHis-tTDH1]*	For western blot and RT-PCR of LjBAS
MD-BA5-t	BY4741, *[pCPS1ULA-BA5-t: CEN6/ARS4-ura3, pGAL1-EtBAS(Sc)-3xFLAG-6xHis-tTDH1]*	For western blot and RT-PCR of EtBAS
MD-BA6-t	BY4741, *[pCPS1ULA-BA6-t: CEN6/ARS4-ura3, pGAL1-PtBAS(Sc)-3xFLAG-6xHis-tTDH1]*	For western blot and RT-PCR of PtBAS
MD-OA1-t	BY4741, *[pMGSEULS-OA1-t: CEN6/ARS4-ura3, pGAL1-GgBAS(Sc)-tTDH1--pGAL10v1-CYP716A12(Sc)-3xFLAG-6xHis-tENO2--pGAL3-AtATR2(Sc)-3xFLAG-6xHis-tPGK1]*	For western blot of CYP716A12 and ATR2
MD-OA2-t	BY4741, *[pMGSEULS-OA2-t: CEN6/ARS4-ura3, pGAL1-GgBAS(Sc)-tTDH1--pGAL10v1-CYP716A15(Sc)-3xFLAG-6xHis-tENO2--pGAL3-AtATR2(Sc)-3xFLAG-6xHis-tPGK1]*	For western blot of CYP716A15 and ATR2

MD-OA3-t	BY4741, *[pMGSEULS-OA3-t: CEN6/ARS4-ura3, pGAL1-GgBAS(Sc)-tTDH1--pGAL10v1-CYP716AL1(Sc)-3xFLAG-6xHis-tENO2--pGAL3-AtATR2(Sc)-3xFLAG-6xHis-tPGK1]*	For western blot of CYP716AL1 and ATR2
MD-OA4-t	BY4741, *[pMGSEULS-OA4-t: CEN6/ARS4-ura3, pGAL1-GgBAS(Sc)-tTDH1--pGAL10v1-CYP716A52v2(Sc)-3xFLAG-6xHis-tENO2--pGAL3-AtATR2(Sc)-3xFLAG-6xHis-tPGK1]*	For western blot of CYP716A52v2 and ATR2
MD-716A12-tag	BY4741, *[pCPS1ULA-CYP716A12-tag: CEN6/ARS4-ura3, PGAL1-CYP716A12 (synth)-3xFLAG-6xHis-TENO2]*	For RT-PCR of CYP716A12
MD-716A15-tag	BY4741, *[pCPS1ULA-CYP716A15-tag: CEN6/ARS4-ura3, PGAL1-CYP716A15 (synth)-3xFLAG-6xHis-TENO2]*	For RT-PCR of CYP716A15
MD-716AL1-tag	BY4741, *[pCPS1ULA-CYP716AL1-tag: CEN6/ARS4-ura3, PGAL1-CYP716AL1 (synth)-3xFLAG-6xHis-TENO2]*	For RT-PCR of CYP716AL1
MD-716A52v2-tag	BY4741, *[pCPS1ULA-CYP716A52v2-tag: CEN6/ARS4-ura3, PGAL1-CYP716A52v2 (synth)-3xFLAG-6xHis-TENO2]*	For RT-PCR of CYP716A52v2
MD-AtATR2-tag	BY4741, *[pCP1EULA-AtATR2-tag: CEN6/ARS4-ura3, PGAL3-AtATR2 (synth)-3xFLAG-6xHis-TPGK1]*	For RT-PCR of ATR2
MD3	BY4741, *[pCPS1ULA-BA5: CEN6/ARS4-ura3, pGAL1-EtBAS (synth)-tTDH1], [pCP23LLA-UGT1: CEN6/ARS4-leu2, pGAL1-UGT73C10(Sc)-tADH1]*	UGT73C10 (+EtBAS)
MD5	BY4741, *[pCPS1ULA-BA5: CEN6/ARS4-ura3, pGAL1-EtBAS (synth)-tTDH1], [pCP45LLA-UGT3: CEN6/ARS4-leu2, pGAL7-UGT73K1(Sc)-tPGK1]*	UGT73K1 (+EtBAS)
MD6	BY4741, *[pCPS1ULA-BA5: CEN6/ARS4-ura3, pGAL1-EtBAS (synth)-tTDH1], [pCP45LLA-UGT4: CEN6/ARS4-leu2, pGAL7-UGT71G1(Sc)-tPGK1]*	UGT71G1 (+EtBAS)
MD7	BY4741, *[pCPS1ULA-BA5: CEN6/ARS4-ura3, pGAL1-EtBAS (synth)-tTDH1], [pCP5ELLA-UGT5: CEN6/ARS4-leu2, pGAL10v1-UGT73P2(Sc)-tENO2]*	UGT73P2 (+EtBAS)
MD10	BY4741, *[pCPS1ULA-BA5: CEN6/ARS4-ura3, pGAL1-EtBAS (synth)-tTDH1], [pMGSELLS-UGT1,5: CEN6/ARS4-leu2, pGAL1-UGT73C10(Sc)-tADH1--pGAL10v1-UGT73P2(Sc)-tENO2]*	UGT73C10 + UGT73P2 (+EtBAS)
MD14	BY4741, *[pMGSEULS-OA7: CEN6/ARS4-ura3, pGAL1-EtBAS(Sc)-tTDH1--pGAL10v1-CYP716A12(Sc)-tENO2--pGAL3-AtATR2(Sc)-tPGK1], [pCP23LLA-UGT1: CEN6/ARS4-leu2, pGAL1-UGT73C10(Sc)-tADH1]*	UGT73C10 (+EtBAS + CYP716A12)

MD15	BY4741, *[pMGSEULS-OA7: CEN6/ARS4-ura3, pGAL1-EtBAS(Sc)-tTDH1--pGAL10v1-CYP716A12(Sc)-tENO2--pGAL3-AtATR2(Sc)-tPGK1], [pCP34LLA-UGT2: CEN6/ARS4-leu2, pGAL7-UGT73F3(Sc)-tPGK1]*	UGT73F3 (+EtBAS + CYP716A12)
MD16	BY4741, *[pMGSEULS-OA7: CEN6/ARS4-ura3, pGAL1-EtBAS(Sc)-tTDH1--pGAL10v1-CYP716A12(Sc)-tENO2--pGAL3-AtATR2(Sc)-tPGK1], [pCP45LLA-UGT3: CEN6/ARS4-leu2, pGAL7-UGT73K1(Sc)-tPGK1]*	UGT73K1 (+EtBAS + CYP716A12)
MD17	BY4741, *[pMGSEULS-OA7: CEN6/ARS4-ura3, pGAL1-EtBAS(Sc)-tTDH1--pGAL10v1-CYP716A12(Sc)-tENO2--pGAL3-AtATR2(Sc)-tPGK1], [pCP45LLA-UGT4: CEN6/ARS4-leu2, pGAL7-UGT71G1(Sc)-tPGK1]*	UGT71G1 (+EtBAS + CYP716A12)
MD18	*BY4741, [pMGSEULS-OA7: CEN6/ARS4-ura3, pGAL1-EtBAS(Sc)-tTDH1--pGAL10v1-CYP716A12(Sc)-tENO2--pGAL3-AtATR2(Sc)-tPGK1], [pCP5ELLA-UGT5: CEN6/ARS4-leu2, pGAL10v1-UGT73P2(Sc)-tENO2]*	UGT73P2 (+EtBAS + CYP716A12)
MD21	BY4741, *[pMGSEULS-OA7: CEN6/ARS4-ura3, pGAL1-EtBAS(Sc)-tTDH1--pGAL10v1-CYP716A12(Sc)-tENO2--pGAL3-AtATR2(Sc)-tPGK1], [pMGSELLS-UGT1,5: CEN6/ARS4-leu2, pGAL1-UGT73C10(Sc)-tADH1--pGAL10v1-UGT73P2(Sc)-tENO2]*	UGT73C10 + UGT73P2 (+EtBAS + CYP716A12)
MD-N1	BY4741, *[pMGSEULS: CEN6/ARS4-ura3, empty vector]*	Empty URA3 vector
MD-N2	BY4741, *[pMGSEULS: CEN6/ARS4-ura3, empty vector; pMGSELLS: CEN6/ARS4-leu2, empty vector]*	Empty URA3 and LEU2 vectors

Genes with the suffix “(Sc)” or “(synth)” were synthesised with codon-optimisation for *S. cerevisiae*.

2.4 List of primers

All primers were provided by Integrated DNA Technologies (IDT).

Primer	Sequence	Notes
PGAL1 fwd	GCATCGTCTCATCGGTCTCAAACGGC TATATTGAAGTACGGATTAGAAGCC	Used to clone pGAL1 from pESC-URA into pYTK001
PGAL1 rev	ATGCCGTCTCAGGTCTCACATAGTTTT TTCTCCTTGACGTTAAAGTATAGAGG	Used to clone pGAL1 from pESC-URA into pYTK001
PGAL10 fwd 1	GCATCGTCTCATCGGTCTCAAACGCTT TATTGTTCGGAGCAGTGCGG	Used to clone pGAL10 from pESC-URA into pYTK001
PGAL10 rev	ATGCCGTCTCAGGTCTCACATAGTTTT GAATTTTCAAAAATTCTTACTTTTTTT TTGGATGGACGC	Used to clone pGAL10 from pESC-URA into pYTK001
(MoClo)-PGAL3 fwd	GCATCGTCTCATCGGTCTCAAACG TTGCTAGCCTTTTCTCGGTCTTG	Used to clone pGAL3 from BY4741 gDNA into pYTK001
(MoClo)-PGAL3 rev	ATGCCGTCTCAGGTCTCACATAGTT TTACTATGTGTTGCCCTACCTTTTT AC	Used to clone pGAL3 from BY4741 gDNA into pYTK001
(MoClo)-PGAL7 fwd	GCATCGTCTCATCGGTCTCAAACGT TTGCCAGCTTACTATCCTTCTTG	Used to clone pGAL7 from BY4741 gDNA into pYTK001
(MoClo)-PGAL7 rev	ATGCCGTCTCAGGTCTCACATAGTT TTTTTTGAGGGAATATTCAACTGTT TTTTTTTATC	Used to clone pGAL7 from BY4741 gDNA into pYTK001
(MoClo)-SAD1 (JIC) fwd	GCATCGTCTCATCGGTCTCATATGTGG AGGCTAACAATAGGTGAGGGC	Used to clone AsBAS into pYTK001
(MoClo)-SAD1 (JIC) rev	ATGCCGTCTCAGGTCTCAGGATTCAGC TCTTAATCGCAAGAAGTCGACGGC	Used to clone AsBAS into pYTK001
dAaBAS fwd 1	GCATCGTCTCATCGGTCTCATATGTGG CGTTTG	Used to make dAaBAS
dAaBAS rev 1	ATCCGTCTCATAGAAACTTGCCAACCA TGATCTTG	Used to make dAaBAS
dAaBAS fwd 2	ATCCGTCTCATCTAATTGCACTGCTGA AGCTTTG	Used to make dAaBAS
dAaBAS rev 2	ATGCCGTCTCAGGTCTCAGGATCCAGT ACCTTTC	Used to make dAaBAS

CqBAS correct fwd 1	GCATCGTCTCATCGGTCTC	Used to generate CqBAS1 without transposon
CqBAS correct rev 1	ATGCCGTCTCAAACATTCTTAAA CAAATGTAGTTCAAGGCAG	Used to generate CqBAS1 without transposon
CqBAS correct fwd 2	GCATCGTCTCATGTTAGGTATA GGTCCTGATGAAGGTGATG	Used to generate CqBAS1 without transposon
CqBAS correct rev 2	ATGCCGTCTCAGGTCTCAG	Used to generate CqBAS1 without transposon
AsBAS (JIC) RT fwd	CATGGACCCTTTCAAGTGTAGATAATG	RT-PCR
AsBAS (JIC) RT rev	TAGTTGAAGTTCAAGGAGGAGTTGAAG	RT-PCR
PgBAS1 (synth) RT fwd	TCTGGCGACTTTAAGTCTATGTATAGG	RT-PCR
PgBAS1 (synth) RT rev	CAATGGAACCCTTCTTCTATATTCAGC	RT-PCR
GgBAS (synth) RT fwd	GAAAGGTTGTATGATTCCGTCAATGTC	RT-PCR
GgBAS (synth) RT rev	GCCCACATTGGATAAATATCTCTGTAC	RT-PCR
LjBAS (synth) RT fwd	CTCAAGAATGGTTGGAATTGTTGAACC	RT-PCR
LjBAS (synth) RT rev	CCCTATACATTGGATAATGCAGCATAC	RT-PCR
EtBAS (synth) RT fwd	GTTGTTTACTGTTCTCTATGATGCCAC	RT-PCR
EtBAS (synth) RT rev	CAATGGAACTCTGTTTCTATATTCGGC	RT-PCR
PtBAS (synth) RT fwd	CTTGGACTTTCTCTGATCAAGATCATG	RT-PCR
PtBAS (synth) RT rev	GTTCTTCATAAAAACGCCAGTGACTTC	RT-PCR
716A12 (synth) RT fwd	GTACTCCATTCAACAAAGCCATTAAGG	RT-PCR

716A12 (synth) RT rev	GACCTTTTCCCACTTAAACCTTTTGAC	RT-PCR
716A15 (synth) RT fwd	GTACTCCATTTCATAGAGCCATCAAAG	RT-PCR
716A15 (synth) RT rev	GATGCAATCTAACTGGCAAACCTTTAG	RT-PCR
716AL1 (synth) RT fwd	GTAAGGCTTCTCCAACTCAAGATATTC	RT-PCR
716AL1 (synth) RT rev	CCATTTGAACCTCTTAACCAGATTGTG	RT-PCR
716A52v2 (synth) RT fwd	CCATTTAACTCCGCTATTAAATCCTCC	RT-PCR
716A52v2 (synth) RT rev	GATTTCCAATCTAGCGTATTCTTTGCC	RT-PCR
AtATR2 (synth) RT fwd	CATAAGGGTGTTTGTTCTACTTGGATG	RT-PCR
AtATR2 (synth) RT rev	ACCAAACATCTCTCAAATACCTACCAG	RT-PCR
oRH711	GTTACGTCGCCTTGGACTTC	Positive control for RT-PCR, binding ACT1; designed by Rebecca Holmes
oRH714	GGTACCACCGGACATAACGA	Positive control for RT-PCR, binding ACT1; designed by Rebecca Holmes

Chapter 3: A GC-MS method to measure triterpenoid production

3.1 Introduction

A major focus of the current work was the generation of yeast strains capable of producing triterpenoids, in particular, β-amyrin and its oxygenated derivatives erythrodiol, oleanolic aldehyde and oleanolic acid (Figure 3.1). These derivatives are generated from β-amyrin by the oxidation of its *C*-28 methyl group to a hydroxyl, aldehyde or carboxyl group, respectively. In addition to these triterpenoids, it was important to monitor the accumulation of ergosterol (Figure 3.1), an essential yeast metabolite that, like β-amyrin, is derived from oxidosqualene. Expression of triterpenoid biosynthetic pathways might therefore affect ergosterol levels by redirecting flux to triterpenoid biosynthesis. Central to this work was a method to monitor triterpenoid production, and gas chromatography-mass spectrometry (GC-MS) was chosen for this purpose. Methods to monitor triterpenoid production, and the chosen GC-MS method, are discussed below.

3.2 Methods to detect triterpenoids

To monitor triterpenoid production it is important to choose an appropriate analytical method. The method of choice will differ depending on the nature of the triterpenoid in question. Two of the most commonly used techniques to detect triterpenoids, either from plant extracts or from heterologous hosts, are GC-MS and liquid chromatography-mass spectrometry (LC-MS). They work by chromatographically separating analytes in a sample, then ionising and fragmenting them as they are eluted into a mass spectrometer. Each analyte gives a unique fragmentation pattern, called a mass spectrum, and by comparing the retention time and mass spectrum with those of an authentic standard, the analyte of interest can be identified. Even where a standard is not available, the identity of an analyte can often be deduced using knowledge of the analyte and its predicted fragmentation pattern, and by comparison with published mass spectra in the literature.

These techniques have high sensitivity and resolution, enabling the detection of small quantities of analyte and the separation of structurally similar compounds. While relatively low throughput, they are amenable to automation, with modern instruments capable of automating sample injection using an autosampler, and some instruments

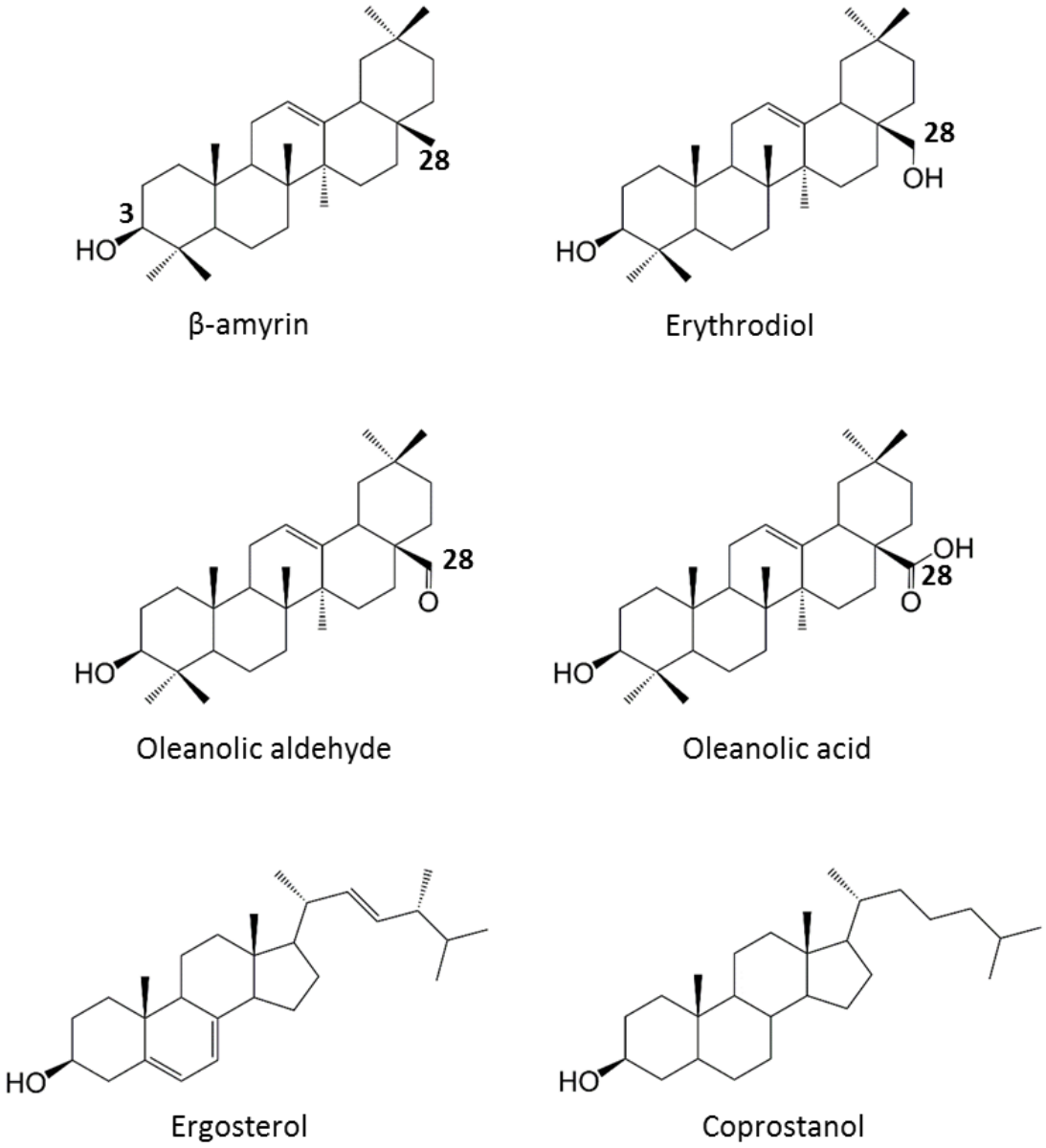


Figure 3.1. Chemical structures of β-amyrin, its *C*-28 derivatives erythrodiol, oleanolic aldehyde and oleanolic acid, and the sterols ergosterol and coprostanol.

able to perform sample preparation (extraction and derivatisation of metabolites). Other techniques such as thin layer chromatography (TLC) and high-performance liquid chromatography (HPLC) are also used, but rely only on the retention time to identify compounds, as they are not coupled to a mass spectrometer. While TLC is fast and inexpensive, it is far less sensitive and is unable to provide absolute identification of an analyte. HPLC combines high sensitivity with a greater throughput than GC-MS or LC-MS, and is commonly used for screening, but again does not have the same accuracy in analyte identification as mass spectrometry-based methods.

The choice of analytical method depends partly on the nature of the analytes in question. GC-MS requires the analytes to be volatile and thermostable, as they are first vaporised at high temperatures (e.g. 250 °C) into the gas phase for chromatographic separation. To increase their volatility, analytes are often treated with a derivatising reagent that reacts with polar moieties (e.g. hydroxyl groups), preventing them from forming hydrogen-bonds. Nevertheless, highly polar and non-volatile analytes still present challenges for GC-MS analysis. By contrast, LC-MS uses a liquid mobile phase and does not require the analytes to be volatile or thermostable, although they must be soluble in the mobile phase. The ionisation method also differs between GC-MS and LC-MS. GC-MS typically uses hard ionisation methods such as electron ionisation (EI), which causes extensive fragmentation of the analyte and produces more complex mass spectra, facilitating identification of the analyte. By contrast, LC-MS uses soft ionisation methods such as electrospray ionisation (ESI), resulting in less fragmentation and therefore providing less information about the analyte. To overcome this limitation, liquid chromatography-tandem mass spectrometry (LC-MS/MS) can be performed. Here, selected ions are fragmented again, giving additional mass spectra for each analyte, facilitating their identification.

Triterpenes are often subject to modification by tailoring enzymes, the most common modifications being oxygenations by cytochromes P450 and glycosylations by UGTs (Thimmappa et al., 2014). These increase their polarity and are responsible for the diverse bioactivities observed in this family of metabolites (Thimmappa et al., 2014). Non-glycosylated triterpenoids that have undergone relatively few modifications by tailoring enzymes are hydrophobic, as they possess few polar moieties. For example, β-amyrin has a single hydroxyl group and is otherwise non-polar. Such triterpenes are well suited to GC-MS analysis, as they are thermostable and can be easily vaporised

under high temperatures. Prior to analysis, triterpenoids are typically derivatised by trimethylsilylation, where a trimethylsilyl (TMS) group reacts with the reactive groups (e.g. hydroxyl or carboxyl) of the triterpenoid, preventing hydrogen-bonding and increasing volatility. The hard ionisation method used during GC-MS analysis results in extensive fragmentation of the triterpenoid. A key fragmentation that occurs for many triterpenoids is the retro Diels-Alder (rDA) reaction, which results in cleavage of the triterpenoid across the C-ring to produce two fragments (Figure 3.2) (Budzikiewicz et al., 1963). Further fragmentation generates an information-rich mass spectrum that facilitates compound identification.

Triterpenoids that have been oxygenated or glycosylated by tailoring enzymes contain an increased number of reactive groups, and the hydrogen-bonds they form reduce volatility, complicating GC-MS analysis. Although derivatisation can be used to increase volatility, oxidised triterpenoids have more sites that must react with the derivatising agent. For highly oxidised triterpenoids that possess many reactive groups, complete derivatisation can be difficult, prohibiting analysis by GC-MS. Furthermore, glycosylated triterpenoids (saponins) are not thermostable, and decompose to lose their sugar moieties under high temperatures. As such, saponins and highly oxidised triterpenoids are not suitable for GC-MS analysis, and LC-MS must be performed instead. The soft ionisation techniques employed by LC-MS instruments result in far less fragmentation of the triterpenoid compared with GC-MS, and notably do not cause rDA fragmentation. The resulting mass spectra therefore provide less information about the triterpenoid. As such, techniques like LC-MS/MS are often employed to allow identification of the triterpenoid or saponin.

In summary, GC-MS is best suited for the analysis of hydrophobic non-glycosylated triterpenoids, while saponins and highly oxidised triterpenoids are better analysed by LC-MS. Generally, GC-MS is preferable because the more complex fragmentation patterns facilitate compound identification. However, the high polarities and thermolability of saponins necessitates the use of LC-MS for their analysis. As such, GC-MS was chosen to monitor the production of β-amyrin, its *C*-28 derivatives and ergosterol, which are all largely hydrophobic (Figure 3.1), while LC-MS was used to monitor saponin production.

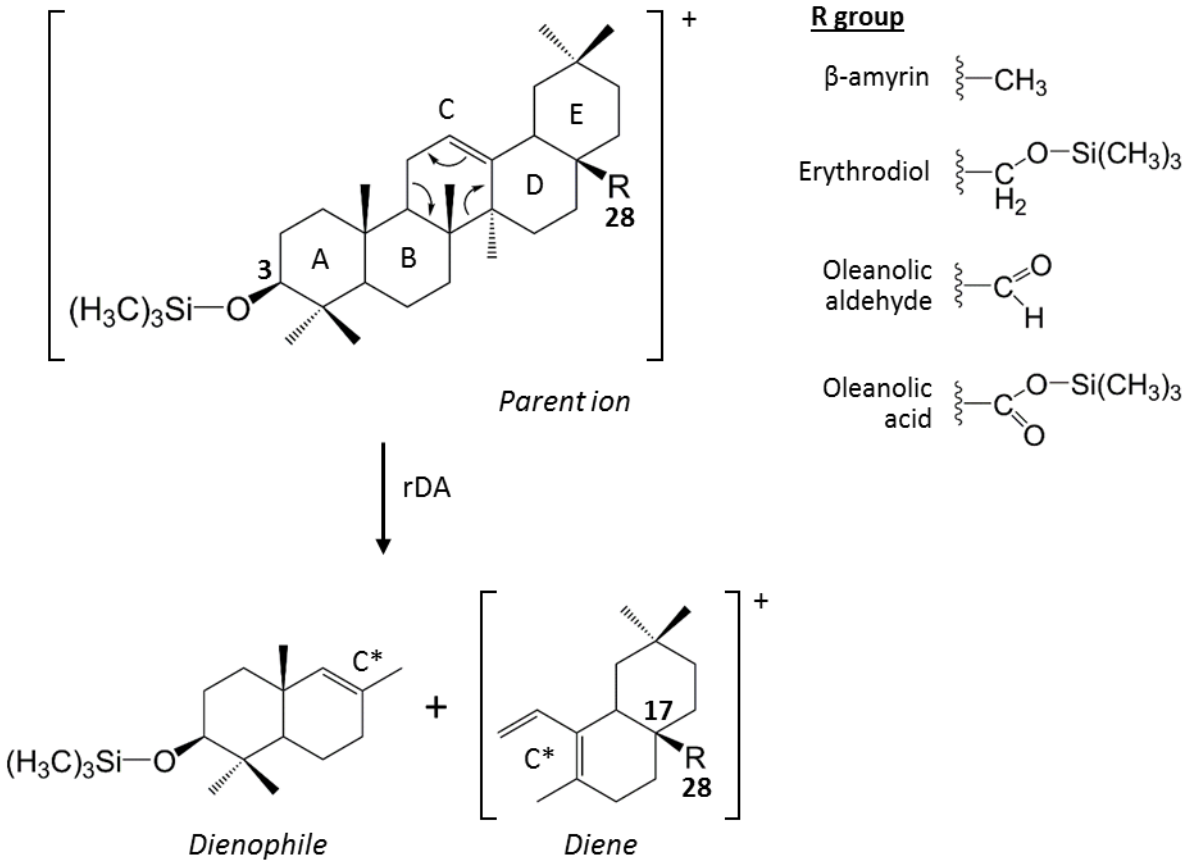


Figure 3.2. Retro Diels-Alder fragmentation of β-amyrin and its C-28 oxygenated derivatives under electron ionisation (EI). The rings are labelled, as are carbons 3 and 28. C* refers to the fragmented C-ring.

3.3 Interpreting triterpenoid chromatograms and mass spectra generated by GC-MS

Following GC-MS analysis, analytes can be identified based on their retention times and mass spectra. The retention time is determined by how strongly the analyte interacts with the stationary phase, which is a liquid adsorbed to the column. Polar analytes interact more strongly with the column and so tend to elute later (i.e. have longer retention times), while hydrophobic analytes elute earlier. Thus, in the analysis of triterpenoids, simple triterpenes such as β-amyrin should elute earlier, while their oxidised derivatives should elute later (Figure 3.3).

As stated above, GC-MS uses hard ionisation techniques such as electron ionisation (EI), which cause extensive fragmentation of the analyte. EI was used in the current study and causes the analyte to lose an electron, generating a radical cation called the parent ion. This can be identified in the mass spectrum as an ion with a mass-to-charge ratio (m/z) that is approximately the same as the molecular weight of the analyte. The parent ion can then fragment further, producing a characteristic mass spectrum. Each fragmentation produces a neutral and a charged fragment, the latter being called a fragment ion. For analytes that fragment easily, the parent ion is often in low abundance compared to the fragment ions. Only ions can be detected by the mass spectrometer, so the neutral fragments are not observed in the mass spectrum.

Trimethylsilylation of the triterpenoids increases their molecular weight by 72 Da for each TMS group added. The parent ion therefore has a correspondingly higher m/z value. For example, β-amyrin has a single hydroxyl group and so its molecular weight increases from 426 Da to 498 Da after derivatisation. The parent ion thus has an m/z of 498 (Table 3.1). Triterpenoids with multiple reactive groups, such as oleanolic acid, gain multiple TMS groups, resulting in higher m/z values for the parent ion.

Pentacyclic triterpenoids with a *C*-12/*C*-13 double bond, such as β-amyrin and its derivatives, undergo rDA fragmentation after ionisation (Figure 3.2). Here, the C-ring is cleaved to produce two fragments: the dienophile, consisting of the ABC* rings, and the diene, consisting of the C*DE rings (where C* means a part of ring C) (Budzikiewicz et al., 1963). For the most part, the diene retains the charge, while the dienophile is neutral and is therefore not observed in the mass spectrum (Budzikiewicz et al., 1963). Additional fragmentations include the loss of methyl groups from the parent or fragment ions, and the loss of TMS groups, which can

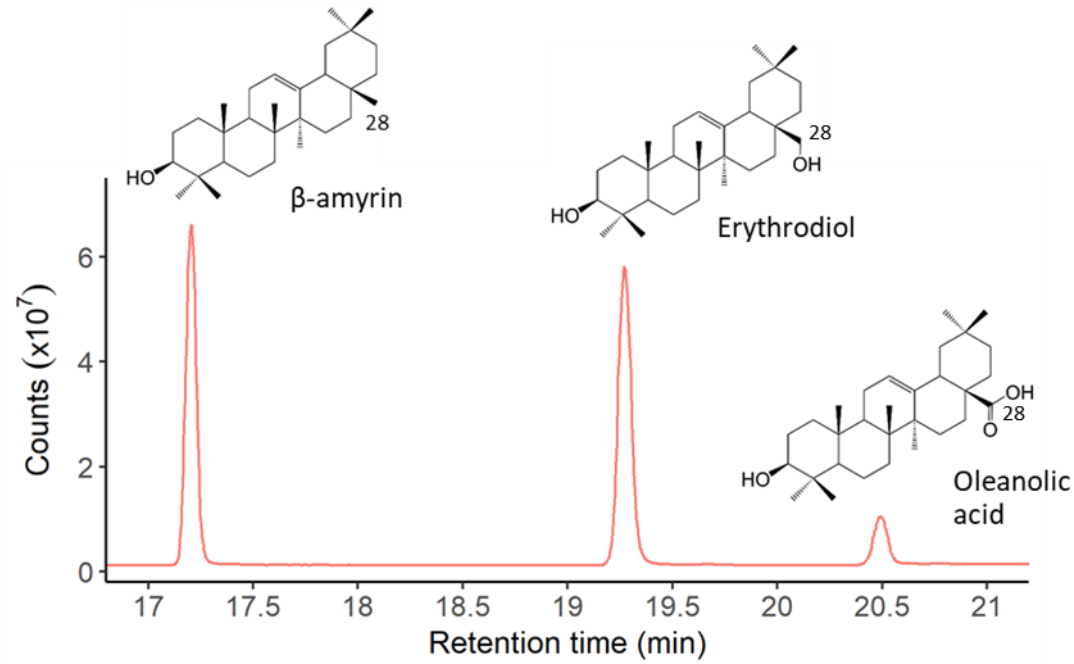


Figure 3.3. Effect of triterpenoid polarity on retention time. Shown is a total ion chromatogram (TIC) of β-amyrin, erythrodiol and oleanolic acid standards generated by GC-MS. β-amyrin, where *C*-28 is a methyl group (labelled) elutes first, followed by erythrodiol (hydroxyl at *C*-28) and oleanolic acid (carboxyl at *C*-28).

be accompanied by the loss of the hydroxyl and/or carboxyl groups to which they are conjugated. A prominent peak with an m/z of 73 is present in all mass spectra, and corresponds to $[TMS]^{\bullet+}$ ions that have been lost from the triterpenoid. Theoretical triterpenoid fragmentation patterns are shown in Table 3.1.

For β-amyrin, the diene is the most abundant ion and has an m/z of 218 (Figure 3.4 A). It is accompanied by two additional fragment ions with m/z values of 203 and 189. The former likely derives from the loss of a methyl group (15 Da) from the diene, while the latter may be due to the loss of the *C*-28 methyl, the *C*-17 carbon (of which the *C*-28 methyl is a substituent) and two accompanying hydrogens (29 Da total) (Budzikiewicz et al., 1963). The dienophile also has an m/z of 189 but as stated above, this is typically a neutral fragment so is unlikely to contribute much signal to the mass spectrum (Budzikiewicz et al., 1963). In addition, loss of TMSOH (at the *C*-3 hydroxyl position; 90 Da) from the parent ion (m/z 498) generates a fragment ion with an m/z of 408, and further loss of a methyl group gives a fragment ion with an m/z of 393. However, these ions are in low abundance, and the mass spectrum is dominated by the diene and its fragment ions (m/z 218, 203 and 189). The diene does not contain a TMS group, so loss of TMSOH is not observed from the rDA fragments.

Because erythrodiol possesses a hydroxyl group at *C*-28, its diene contains a TMS group and therefore has an m/z of 306 (Figure 3.4 B). However, this is in very low abundance, with loss of TMSOH (90 Da) from the diene producing a fragment ion with an m/z of 216, which is the most abundant ion. This fragment ion is characteristic of β-amyrin derivatives containing a single hydroxyl group in their diene fragment (Yasumoto et al., 2016). Loss of the entire *C*-28 group ($TMSOCH_2$; 103 Da) gives an m/z of 203, and further loss of the *C*-17 carbon and two accompanying hydrogens (14 Da) gives an m/z of 189. These are equivalent to the 203 and 189 m/z fragment ions for β-amyrin. A peak of m/z 188 with a similar abundance to the m/z 189 peak is possibly explained by loss of $TMSOCH_2$ and a methyl group (103 and 15 Da) from the diene. The erythrodiol parent ion (m/z 586) and a fragment ion corresponding to the loss of a methyl group from the parent (m/z 571) are in low abundance. However, an ion with an m/z of 496, corresponding to the loss of TMSOH from the parent ion, is far more abundant. This contrasts with β-amyrin, whose abundant fragment ions appear to be derived from the rDA reaction. In addition, low abundance fragment ions

Table 3.1 Theoretical triterpenoid fragmentation patterns under electron ionisation.

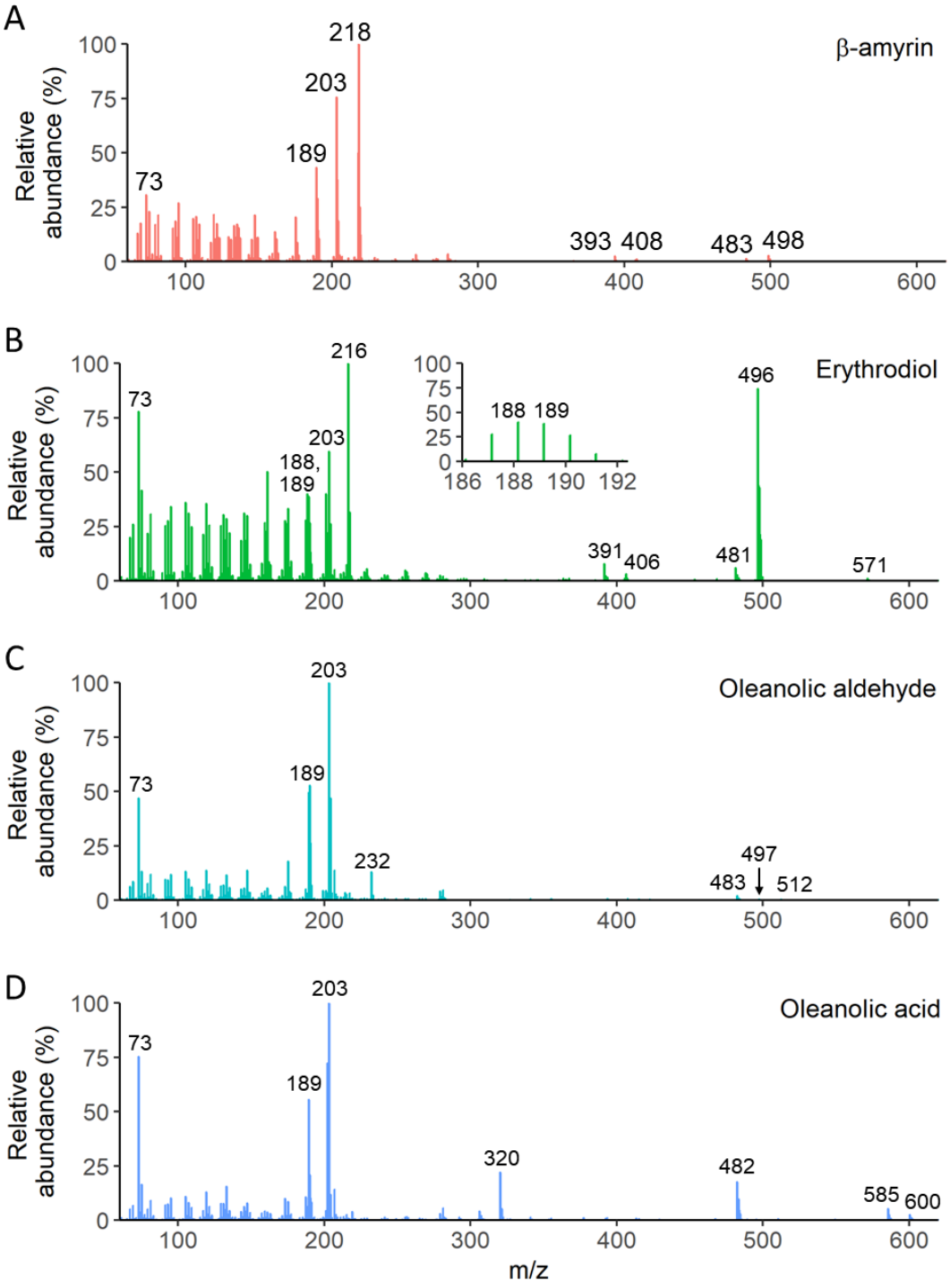
Triterpenoid	Molecular weight (Da)	Ions (m/z)
β-amyrin	426	498, parent ion (M^+) (+TMS, -H) 483, M^+ -CH_3 408, M^+ -TMSOH 393, M^+ -TMSOH, -CH_3 218, C*DE rings of M^+ (diene) 203, C*DE rings of M^+ -CH_3 189, C*DE rings of M^+ -CH_3 at C28, -C at C17, -2H
Erythrodiol	442	586, parent ion (M^+) (+ 2TMS, - 2H) 571, M^+ -CH_3 496, M^+ -TMSOH 481, M^+ -TMSOH, -CH_3 406, M^+ -2 TMSOH 391, M^+ -2 TMSOH, -CH_3 306, C*DE rings of M^+ (diene) 216, C*DE rings of M^+ -TMSOH 203, C*DE rings of M^+ -$TMSOCH_2$ 189, C*DE rings of M^+ -$TMSOCH_2$, -C at C17, -2H 188, C*DE rings of M^+ -$TMSOCH_2$, -CH_3
Oleanolic aldehyde	440	512, parent ion (M^+) (+TMS, -H) 497, M^+ -CH_3 232, C*DE rings of M^+ (diene) 203, C*DE rings of M^+ -CHO 189, C*DE rings of M^+ -CHO, -C at C17, -2H
Oleanolic acid	456	600, parent ion (M^+) (+2TMS, -2H) 585, M^+ -CH_3 482, M^+ -TMSCOOH 320, C*DE rings of M^+ (diene) 203, C*DE rings of M^+ -TMSCOO 202, C*DE rings of M^+ -TMSCOOH 189, C*DE rings of M^+ -TMSCOOH, -C at C17, -2H
Ergosterol	396	468, parent ion (M^+) (+TMS, -H) 453, M^+ -CH_3 378, M^+ -TMSOH 363, M^+ -TMSOH, -CH_3 337, A*BCD rings of M^+ 253, M^+ -TMSOH, -C_9H_{17} 131, A* ring of M^+
Coprostanol	388	460, parent ion (M^+) (+TMS, -H) 445, M^+ -CH_3 370, M^+ -TMSOH 355, M^+ -TMSOH, -CH_3

with m/z values of 481 and 406 can be observed. The former may be explained by the loss of a methyl group from the m/z 496 fragment, and the latter by the loss of two TMSOH groups from the parent ion (m/z 586 – 180). A peak with an m/z of 391 is consistent with the loss of a methyl group from the m/z 406 fragment ion.

Oleanolic aldehyde contains an aldehyde at the *C*-28 position, and this group is not trimethylsilylated because it lacks a reactive oxygen. The oleanolic aldehyde diene therefore lacks a TMS group and has an m/z of 232 (Figure 3.4 C). The diene gives a small peak, while fragment ions with m/z values of 203 and 189, which are equivalent to those from β-amyrin and erythrodiol, are more abundant. The parent ion (m/z 512) is in very low abundance, as is a fragment corresponding to the loss of a methyl from the parent ion (m/z 497). A small peak with an m/z of 483 was also observed, and may result from the loss of the aldehyde group (29 Da) from the parent ion. Similar to β-amyrin, the low abundance of ions that do not derive from rDA fragmentation indicates that this reaction goes almost to completion.

Oleanolic acid has a carboxyl at *C*-28 and its diene therefore contains a TMS group, giving it an m/z of 320 (Figure 3.4 D). The oleanolic acid diene is more abundant than for erythrodiol and has a noticeable peak. However, the most abundant ion has an m/z of 203, and corresponds to the loss of TMSCOO (i.e. the TMS group and the carboxyl group to which it is attached) from the diene. A slightly less abundant peak with an m/z of 202 is possibly due to the loss of an additional hydrogen atom (i.e. loss of TMSCOOH), while a fragment ion with an m/z of 189 is equivalent to those observed for β-amyrin, erythrodiol and oleanolic aldehyde. The parent ion (m/z 600) is in low abundance, and fragment ions corresponding to the loss of a methyl or a TMSCOOH group from the parent ion (m/z 585 and 482, respectively) have stronger signals.

In contrast to the above triterpenoids, ergosterol does not have a *C*-12/C-13 double bond (Figure 3.1) and does not appear to undergo rDA fragmentation (Figure 3.4 E). Its parent ion is readily apparent (m/z 468), as are fragment ions corresponding to the loss of TMSOH from the parent ion (m/z 378) and to the additional loss of a methyl group (m/z 363). A weak peak with an m/z of 453 is also observed, and may result from the loss of a methyl group from the parent ion. An abundant fragment ion with an m/z of 337 may result from fragmentation at the A ring, resulting in the loss of 131 Da from the parent ion (468 – 131 = 337) (Kenny and Wetzel, 1995). Indeed, a


A
β-amyrin
Relative abundance (%)
218
203
189
73
393 408
483 498
B
Erythrodiol
216
73
203
188, 189
188 189
496
391 406
481
571
C
Oleanolic aldehyde
203
73
189
232
497
483 512
D
Oleanolic acid
203
73
189
320
482
585 600
m/z

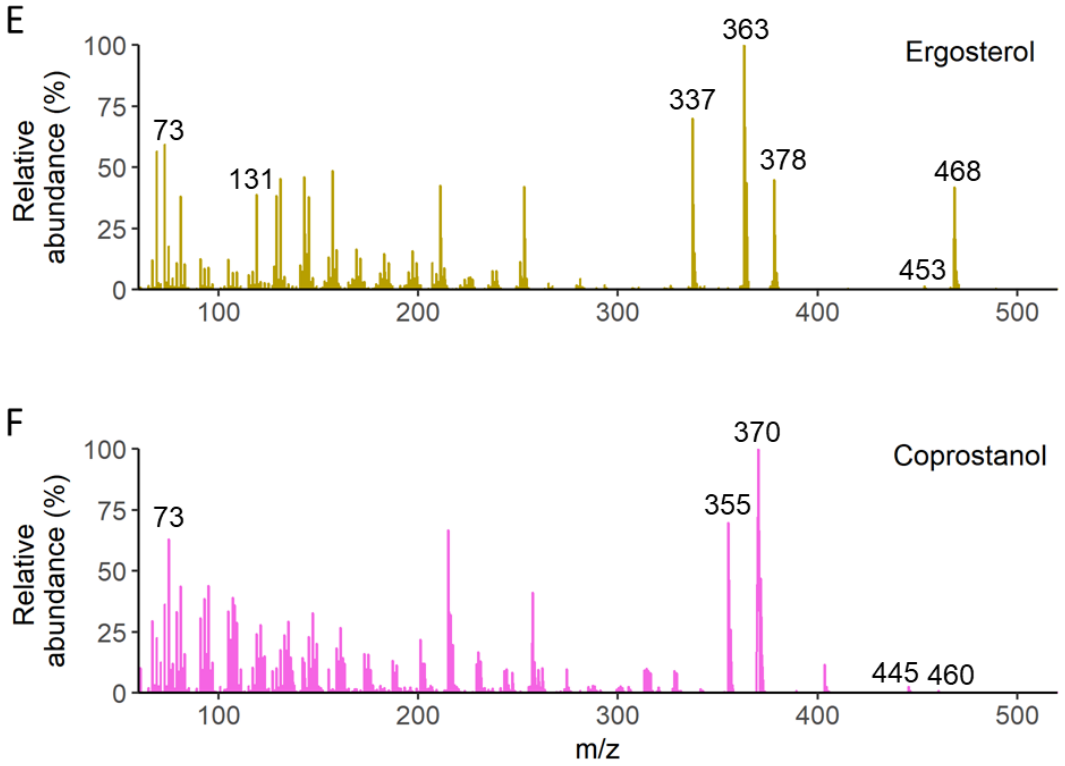


Figure 3.4. Mass spectra of β-amyrin (A), erythrodiol (B), oleanolic aldehyde (C), oleanolic acid (D), ergosterol (E) and coprostanol (F). The mass spectra for all analytes but oleanolic aldehyde were obtained by analysing authentic standards by GC-MS (1 µl of 50 ng/µl derivatised standard injected with a 10:1 split). For oleanolic aldehyde, the mass spectrum of a putative oleanolic aldehyde peak from strain MD-OA15 (expressing an oleanolic acid biosynthesis pathway) is shown. This was obtained by injecting 1 µl of derivatised metabolite extract at a 10:1 split. The mass spectrum is consistent with that reported for an authentic oleanolic aldehyde standard in the literature (see Fig. S3 of Khakimov *et al.*, 2015). The erythrodiol parent ion (m/z 586) and diene (m/z 306) are not labelled as they were not detected.

fragment ion with an m/z of 131 is also observed, and likely corresponds to the A ring fragment (denoted A*). Finally, a fragment ion with an m/z of 253 may result from the loss of the aliphatic hydrocarbon chain (125 Da) from the m/z 378 fragment ion.

Coprostanol, a sterol that is not produced by yeast and was used as an internal standard for all GC-MS analyses (see below), also does not appear to undergo rDA fragmentation. It has a similar structure to ergosterol, and also lacks a *C*-12/C-13 double bond (Figure 3.1). However, in contrast with ergosterol, its parent ion (m/z 460) is in very low abundance (Figure 3.4 F). In slightly greater abundance is a fragment ion with m/z 445, which may result from the loss of a methyl group from the parent ion. Similar to ergosterol, it has strong peaks resulting from the loss of TMSOH from the parent ion (m/z 370) and the further loss of a methyl group (m/z 355).

3.4 A GC-MS method to detect triterpenoids

GC-MS was chosen to monitor the production of non-glycosylated triterpenoids in yeast. Indeed, GC-MS is the method of choice to measure triterpenoid production in engineered yeast strains (Fukushima et al., 2013; Moses et al., 2014b; Moses et al., 2014c; Arendt et al., 2017; Miettinen et al., 2017; Zhang et al., 2015; Seki et al., 2011; Zhao et al., 2018; Zhu et al., 2018; Dai et al., 2014; Kirby et al., 2008). Metabolites are typically extracted one to three times with hexane or ethyl acetate, which favour extraction of the largely hydrophobic triterpenoids, and the crude extracts are then trimethylsilylated before being injected onto the column for GC-MS analysis (Fukushima et al., 2013; Moses et al., 2014b; Moses et al., 2014c; Arendt et al., 2017; Miettinen et al., 2017; Zhang et al., 2015; Seki et al., 2011; Zhao et al., 2018; Zhu et al., 2018). The trimethylsilylation reaction is usually performed using *N*-Methyl-*N*-(trimethylsilyl)trifluoroacetamide (MSTFA), but incubation times and temperatures vary between studies (Fukushima et al., 2013; Moses et al., 2014b; Moses et al., 2014c; Arendt et al., 2017; Miettinen et al., 2017; Zhang et al., 2015; Seki et al., 2011; Zhao et al., 2018). In addition, one study instead used *N*,*O*-Bis(trimethylsilyl) trifluoroacetamide (BSTFA) (Zhu et al., 2018); it was not clear if the use of BSTFA provided any advantage over MSTFA.

Similar GC columns and GC-MS running parameters were also used in these studies. The columns were non-polar with similar stationary phases suitable for the separation of triterpenoids: DB-1/DB-1ms (100% dimethylpolysiloxane stationary phase) (Seki et al., 2011; Fukushima et al., 2013), DB-5ms (phenyl arylene polymer, equivalent to

(5%-Phenyl)-methylpolysiloxane) (Kirby et al., 2008), HP-5/HP-5ms [(5%-phenyl)-methylpolysiloxane] (Dai et al., 2014; Seki et al., 2011), VF-5ms (5% phenylmethyl) (Moses et al., 2014; Moses et al., 2014c; Arendt et al., 2017; Miettinen et al., 2017) and Rxi-5Sil MS [1,4-bis(dimethylsiloxy)phenylene dimethyl polysiloxane] (Zhao et al., 2018; Zhu et al., 2018). Where reported, ionisation was performed by electron ionisation (EI) (Moses et al., 2014; Moses et al., 2014c; Arendt et al., 2017; Miettinen et al., 2017; Zhao et al., 2018; Dai et al., 2014). Injection temperatures ranged from 250-300 °C and oven temperatures typically started at 80 °C for 1 min, before being ramped to ~ 300 °C at a rate of 20 °C/min and holding at that temperature for 15-30 min (Fukushima et al., 2013; Seki et al., 2011; Zhao et al., 2018; Dai et al., 2014; Zhu et al., 2018). Alternatively, the temperature was ramped to 280 °C, held at 280 °C for 15-40 min, then ramped again to 300-320 °C (Moses et al., 2014b; Moses et al., 2014c; Arendt et al., 2017; Miettinen et al., 2017; Zhang et al., 2015; Kirby et al., 2008). Thus, similar methods to analyse triterpenoids produced by engineered yeast are employed across a range of studies.

A similar method for the analysis of triterpenoids was developed here, with the help of Hannah Florance (University of Edinburgh). Initially, a mixture of standards was analysed to confirm that they could be detected by GC-MS. These were the triterpenoids β-amyrin, erythrodiol and oleanolic acid, and the sterols ergosterol and coprostanol, which are structurally similar to the triterpenoids (Figure 3.1). Coprostanol is not produced by yeast and was included as an internal standard (discussed below). All standards in the mixture were at a concentration of 50 ng/µl, except coprostanol, which was at 0.18 mM (70 ng/µl). The standards were also run individually to allow each peak in the standards mix to be identified. The dissolved standards were evaporated to dryness, derivatised by trimethylsilylation in 50 µl of MSTFA:pyridine (4:1), and 1 µl was injected at a 10:1 split, meaning that only 1/11th of the sample enters the GC column. The inlet was set to 250 °C, and the oven was held at 100 °C for 2 min at the start of the GC program, ramped to 300 °C at a rate of 25 °C per min, then held at 300 °C for 25 min. Analysis of the total ion chromatograms (TICs) revealed distinct and well resolved peaks (Figure 3.5 A) with the expected mass spectra for each standard (Figure 3.4), indicating that this is a suitable method to monitor triterpenoid production. In addition, peak areas varied depending on the analyte (Figure 3.5 A), despite the same concentration of each analyte being injected. This indicates that peak areas alone cannot be used to compare the amounts of

different analytes. However, use of a standard curve can account for such analyte-dependent variations in peak area.

Next, the GC-MS method was applied to analyse a metabolite extract obtained from a yeast strain expressing an oleanolic acid biosynthetic pathway (MD-OA15). This strain would be expected to produce β-amyrin, erythrodiol, and oleanolic acid, as well as oleanolic aldehyde, for which an authentic standard is not available. Cells were initially grown for 24 h in glucose-supplemented media, diluted into galactose-supplemented media to induce gene expression, and grown for 96 h. The cultures were washed, resuspended in ethanolic KOH containing 80 nmol/ml coprostanol (the internal standard), and lysed by incubating at 100 °C for 10 min. The total lysates were extracted twice in 500 µl hexane then twice in 500 µl ethyl acetate to obtain 2 ml of organic extract. 600 µl of the metabolite extract was vacuum evaporated to dryness, derivatised and analysed as described for the standards. This method was based on one provided by Thomas Louveau (Anne Osbourn lab, John Innes Centre). The expected analytes were observed in the TICs, at the same retention times as the corresponding standards (Figure 3.5 B). By contrast, an empty vector control strain (MD-N1) produced ergosterol but no detectable triterpenoids (Figure 3.5 B).

As stated above, an internal standard (IS) was added to the samples at the cell lysis step. Using an appropriate IS helps to account for extraction efficiency across samples and possible manual errors introduced after its addition to the sample, because sample handling variation should affect the peak areas of the IS and analytes equally. Internal standardisation is performed by adding a fixed concentration of IS to each sample and dividing analyte peak areas by that of the IS, giving normalised peak areas. The choice of IS is important, as it should have a similar structure and physicochemical properties to the analytes of interest, to ensure that they are extracted at similar efficiencies in the solvents used. In this way, any sample handling variation should affect the IS and analytes similarly. In addition, the IS should not be present naturally in the sample, so that the IS peak area corresponds only to what was added to the sample. Coprostanol meets these criteria and has a distinct retention time (Figure 3.5 B). By adding coprostanol at the cell lysis step, error introduced during metabolite extraction, derivatisation, and GC-MS analysis can be accounted for.

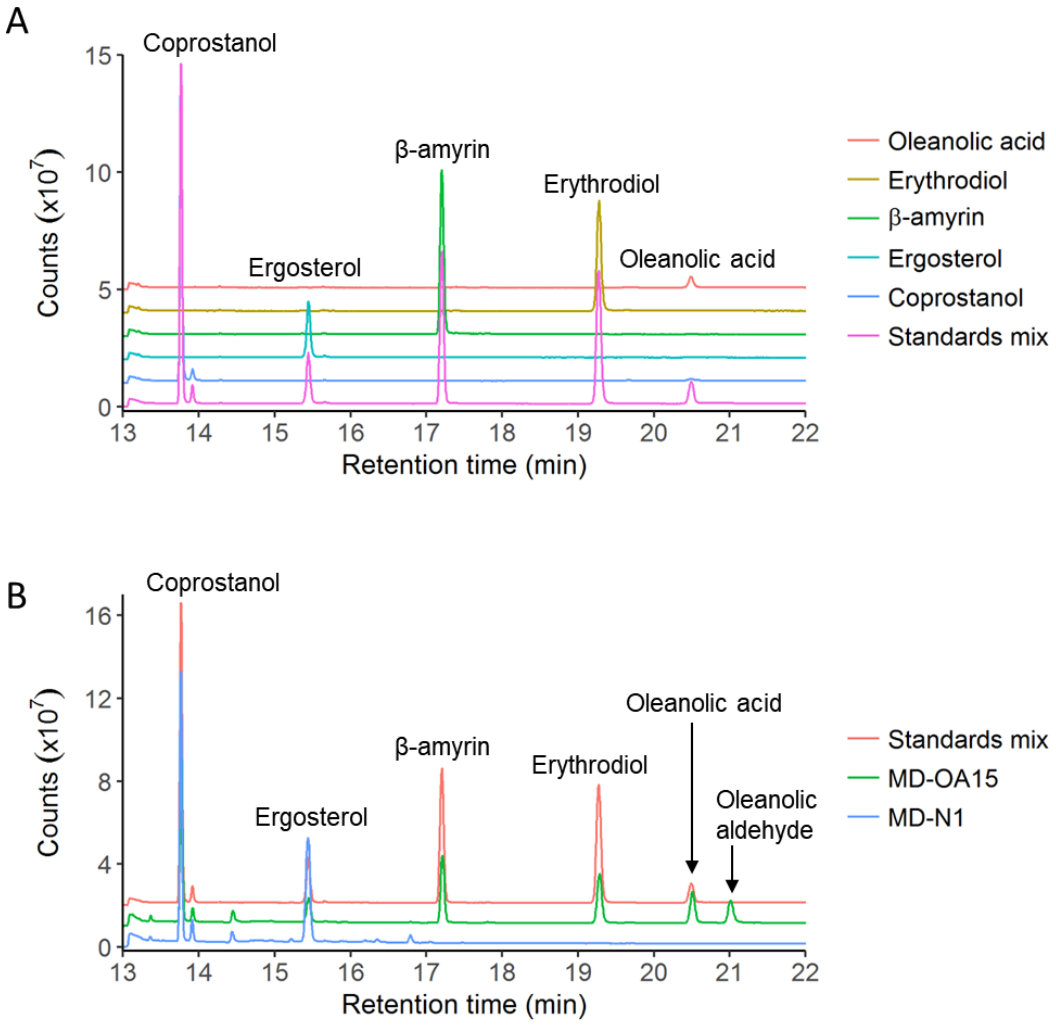


Figure 3.5. Initial analysis of triterpenoids by GC-MS. (A) Overlay of total ion chromatograms (TICs) of the standards mix used to generate the standard curve, alongside individual triterpenoid standards (coprostanol, ergosterol, β-amyrin, erythrodiol and oleanolic acid). **(B)** TIC of MD-OA15 alongside MD-N1 and the standards mix, showing that the expected triterpenoids can be detected using the chosen GC-MS method.

The combination of solvents used for extraction (hexane and ethyl acetate) was chosen to maximise the extraction efficiency for the analytes of interest. In order to compare the levels of different triterpenoids in the sample, it is important that they are extracted at similar efficiencies. Hexane favours the extraction of more hydrophobic compounds, such as β-amyrin, ergosterol and coprostanol, while ethyl acetate efficiently extracts oxidised triterpenoids such as erythrodiol, oleanolic aldehyde and oleanolic acid.

3.5 Quantifying triterpenoid production

GC-MS analysis returns peak areas that are proportional to the concentration of analyte in the sample, but do not represent the actual concentration. To calculate the concentration of an analyte, a standard curve can be constructed by injecting various concentrations of an authentic standard and plotting the resulting peak areas against these known concentrations. The concentration of analyte in the sample is then calculated by reading the analyte peak area off the linear region of the standard curve. Quantification using this method can only be done for analytes for which a standard of known concentration can be obtained. The standard does not have to be pure, but the peaks of any impurities should not overlap with that of the analyte, and the concentration of the standard itself must be known. Nevertheless, high purities are desirable to facilitate analysis. If an IS has been added to the samples, it can be included in the quantification. In this case, normalised peak areas for the samples are calculated by dividing the analyte peak area by that of the IS. The standard curve is constructed by mixing a fixed concentration of IS with various concentrations of the authentic standard and calculating the normalised peak areas. The normalised peak areas from the samples are then read off this standard curve.

Triterpenoid titres were calculated using the internal standard coprostanol and standard curves of authentic standards. A mixture of β-amyrin, erythrodiol, oleanolic acid and ergosterol standards was prepared with each standard at the same concentration (in ng/µl). Various dilutions of this standard mixture were made and coprostanol was added to each to a final concentration of 0.18 mM. The final concentrations of analytes ranged from 0 to 120 ng/µl (increasing in increments of 10 ng/µl). These were made up such that the concentration remained the same after evaporation of the solvent and derivatisation in 50 µl MSTFA:pyridine. The concentration of coprostanol was lower than that added to samples. During sample

preparation, 40 nmol coprostanol is added at the cell lysis step, resulting in a concentration of 0.24 mM coprostanol after derivatisation (assuming 100 % extraction efficiency and complete derivatisation). However, injection of 0.24 mM coprostanol standard resulted in a saturated peak (Figure 3.6), which would cause inaccuracies if used for quantification. By contrast, the coprostanol peak in the samples is not saturated (Figure 3.6). This can be explained by losses of coprostanol during sample extraction, leading to less coprostanol being injected onto the column than would otherwise be expected. Therefore, 0.18 mM coprostanol was used in the standard curve. This gave a similar peak area to the coprostanol in the samples and was not saturated (Figure 3.6).

The standards were derivatised and GC-MS analysis was performed by injecting 1 µl with a 10:1 split. Because the concentration of coprostanol in the standard curve was lower than in the samples, the coprostanol peak areas in the standard curve were multiplied by 4/3 (equal to 0.24 mM / 0.18 mM) to account for this difference, giving a corrected coprostanol peak area. For each concentration of standard, the peak area from the TIC was normalised to the corrected coprostanol peak area and used to construct a standard curve of normalised peak areas against the amount of standard on the column (ng). The amount of standard on the column can be calculated as:

$$Standard\ on\ column\ (ng) = \frac{Concentration\ (ng/\mu l) \times Injection\ volume\ (\mu l)}{Split\ factor}$$

where concentration refers to the concentration of analyte injected, and the split factor is simply the dilution factor due to the split. For a split ratio of A:B, the split factor is A + B. Thus, for a split ratio of 10:1, the split factor is 11. A straight line was fit to the linear region of each standard curve (Figure 3.7), and used to calculate titres. For β-amyrin and erythrodiol, the linear region ranged from 0.91 to 10.9 ng on the column, which would correspond to sample titres of 1.7 to 20.0 mg/L. The linear region for ergosterol and oleanolic acid ranged from 1.82 to 9.09 ng, corresponding to sample titres of 3.3 to 16.7 mg/L.

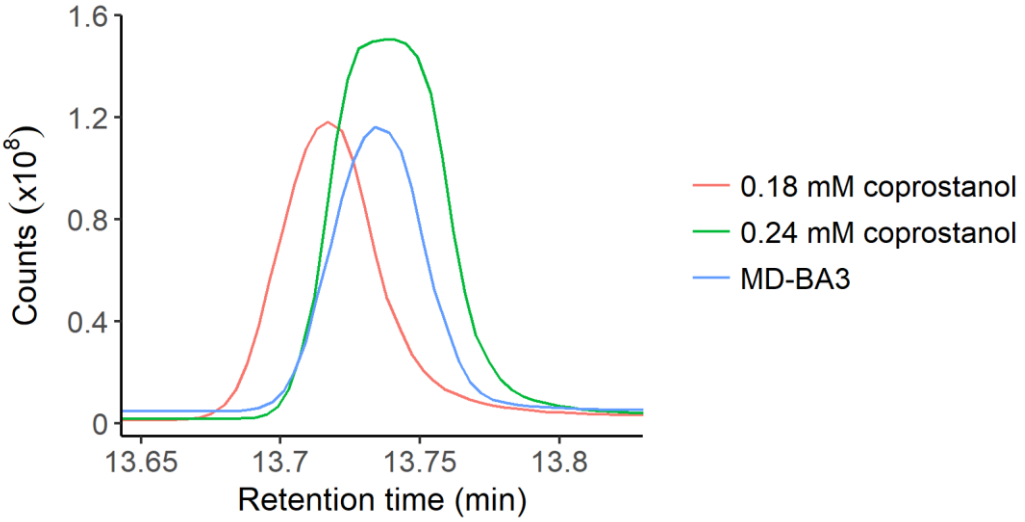


Figure 3.6. Injection of 0.24 mM coprostanol gives a saturated peak. Analysis of TICs shows that injection of 0.24 mM coprostanol (1 µl, 10:1 split) gives a saturated peak, while the peak for 0.18 mM coprostanol is unsaturated and similar in size to the coprostanol in the samples (MD-BA3).

To calculate sample titres, derivatised samples are injected at 1 µl volume with a 10:1 split and normalised peak areas for each analyte are calculated. These are converted to the amount of analyte on the column using the following equation for the appropriate standard curve:

$$Analyte\ on\ column\ (ng) = \frac{PA_{samp} - c}{m}$$

where PA_{samp} is the peak area of the analyte in the sample, and c and m are the intercept and slope of the straight line from the standard curve. The amount of analyte on the column can be converted to a titre (expressed as milligram of analyte per litre of culture, or mg/L) using the following equation:

$$Titre\ (mg/L) = analyte\ on\ column\ (ng) \times split\ factor \times 50 \times \frac{2000}{600 \times 1000}$$

Multiplying the amount of analyte on the column by the split factor gives the amount of analyte injected, and because the injection volume was 1 µl, multiplying by 50 gives the amount of analyte in the 50 µl of derivatisation mixture. This is equal to the amount of analyte in 600 µl crude extract, which is the volume of extract that was evaporated and derivatised. Because a total of 2 ml crude extract was obtained for each sample, we then multiply by 2000/600 to obtain the amount of analyte in the total extract. This is equal to the amount of analyte (ng) extracted from 1 ml of culture, i.e. the titre in ng/ml. Dividing by 1000 converts this to mg/L.

It is important that the sample peak area falls within the linear region of the standard curve, otherwise titres cannot be reliably calculated. For peak areas that fall outside of this range, the sample can be injected at a different split to ensure that the peak area falls within the linear region. In this case, the peak area of the analyte from the new split is normalised to the coprostanol peak area from the same sample injected at the 10:1 split. This is equivalent to changing the concentration of the analyte in the sample while keeping the concentration of coprostanol the same. A sample analyte that falls above the linear region should be injected at a higher split (greater dilution), and an analyte that falls below this region is injected at a lower split.

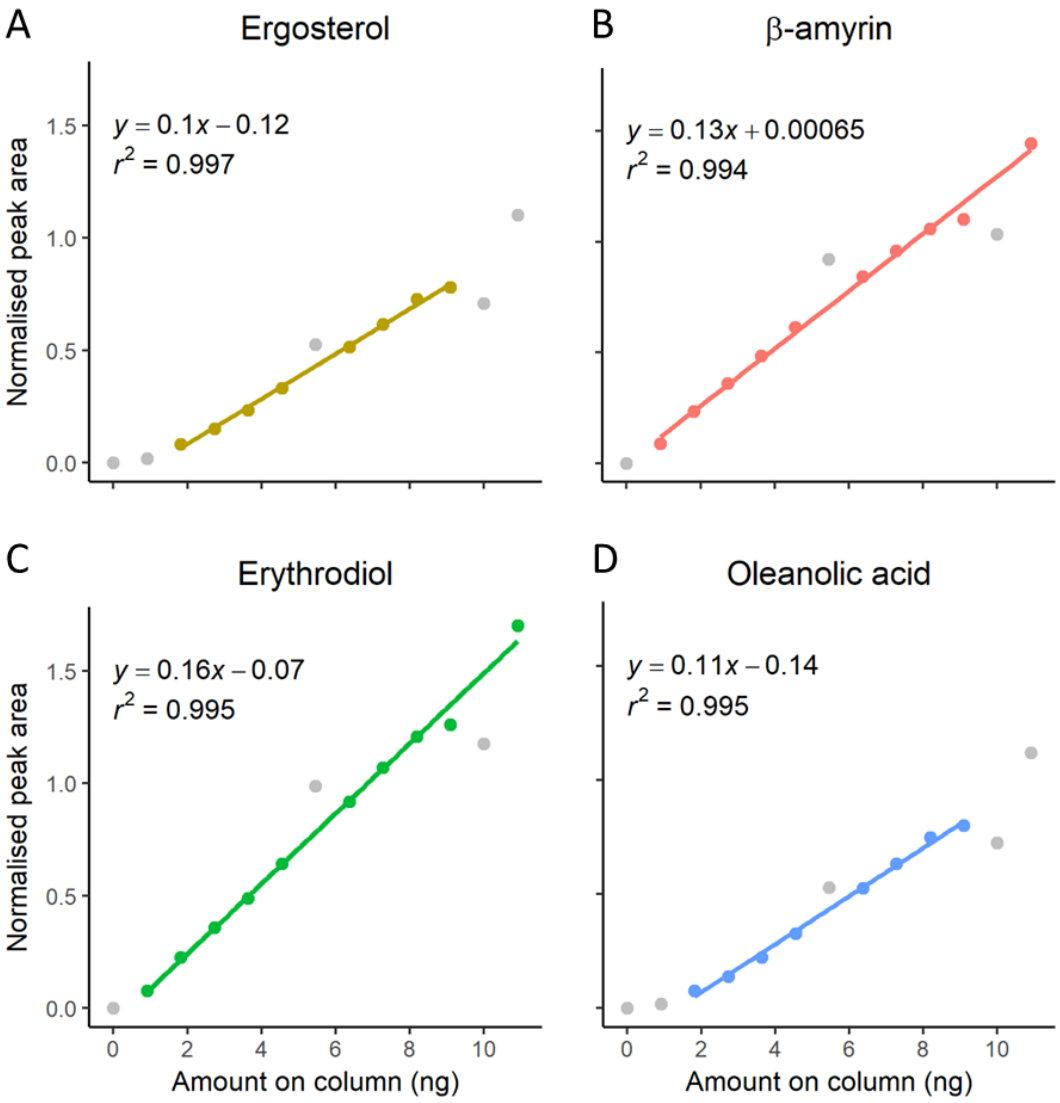


Figure 3.7. Standard curves of ergosterol (A), β-amyrin (B), erythrodiol (C) and oleanolic acid (D). Standard curves were generated by injecting various concentrations of a triterpenoid standards mix (ergosterol, β-amyrin, erythrodiol and oleanolic acid), each with the same concentration of the internal standard coprostanol (0.18 mM). Injections were 1 µl volume at a 10:1 split. Triterpenoid concentrations were from 0 to 120 ng/µl (increasing in increments of 10 ng/µl). The amount of triterpenoid on the column is equal to the concentration divided by 11 (the split factor), so ranges from 0 to 10.9 ng. Each concentration of standards was injected three times sequentially. Shown is the mean ± 1 standard deviation.

3.6 Discussion

A GC-MS method to monitor the production of triterpenoids in engineered yeast has been presented. This method is similar to previous methods for measuring triterpenoid production reported in the literature (Fukushima et al., 2013; Moses et al., 2014b; Moses et al., 2014c; Arendt et al., 2017; Miettinen et al., 2017; Zhang et al., 2015; Seki et al., 2011; Zhao et al., 2018; Zhu et al., 2018; Dai et al., 2014; Kirby et al., 2008). Metabolites were extracted using hexane and ethyl acetate, which are both commonly used to extract triterpenoids (Fukushima et al., 2013; Moses et al., 2014b; Moses et al., 2014c; Arendt et al., 2017; Miettinen et al., 2017; Zhang et al., 2015; Seki et al., 2011; Zhao et al., 2018; Zhu et al., 2018) and, based on their chemical structures, should favour the extraction of the non-polar triterpenoids. The crude extracts were derivatised by trimethylsilylation using a mixture of MSTFA and pyridine and incubating for 30 min at 37 °C. Previous studies have typically derivatised triterpenoid extracts by trimethylsilylation, using either MSTFA alone or a mixture of MSTFA and pyridine (Fukushima et al., 2013; Moses et al., 2014b; Moses et al., 2014c; Arendt et al., 2017; Miettinen et al., 2017; Zhang et al. 2015; Seki et al., 2011; Zhao et al., 2018); where reported, incubations were either at room temperature for 15 min or at 80 °C for 30 min (Moses et al., 2014b; Zhang et al, 2015; Seki et al., 2011).

In addition, GC-MS analysis was performed similarly to that reported in the literature. The column used was the non-polar DB-5ms column, which uses a phenyl arylene polymer stationary phase – equivalent to (5%-phenyl)-methylpolysiloxane – and has been used previously for the analysis of triterpenoids (Kirby et al., 2008; Zhang et al., 2015). Other studies have used similar non-polar columns, for example the HP-5 and HP-5ms columns which use a (5%-phenyl)-methylpolysiloxane stationary phase (Dai et al., 2014; Seki et al., 2011), the DB-1 and DB-1ms columns which have a 100% dimethylpolysiloxane stationary phase (Fukushima et al., 2013; Seki et al., 2011), and the VF-5ms column which has a 5% phenylmethyl stationary phase (Moses et al., 2014b; Moses et al., 2014c; Arendt et al., 2017; Miettinen et al., 2017). As with previous studies (Moses et al., 2014b; Moses et al., 2014c; Arendt et al., 2017; Miettinen et al., 2017; Zhao et al., 2018; Dai et al., 2014), electron ionisation (EI) was used to ionise the metabolites, and the injection temperature was 250 °C (compared with 250-300 °C used in previous studies) (Fukushima et al., 2013; Moses et al., 2014b; Moses et al., 2014c; Arendt et al., 2017; Miettinen et al., 2017; Zhang et al.,

2015; Seki et al., 2011; Zhao et al., 2018; Dai et al., 2014; Zhu et al., 2018). The oven temperatures were also similar to those described previously, starting at 100 °C for 2 min before being ramped to 300 °C at a rate of 25 °C/min then being held at 300 °C for 25 min. Previous studies have used an 80 °C starting temperature for 1 min, followed by a ramp to 290-300 °C at a rate of 20 °C/min then holding at that temperature for 15-30 min (Fukushima et al., 2013; Seki et al., 2011; Zhao et al., 2018; Dai et al., 2014; Zhu et al., 2018).

It is difficult to compare the quantification method described here with those used in previous studies, as the precise method of quantification is often not described in detail (Moses et al., 2014b; Arendt et al., 2017; Kirby et al., 2008; Zhao et al., 2018). In the present method, sample peak areas were first normalised to that of the internal standard (IS) coprostanol before being compared to a standard curve of an authentic triterpenoid standard (also normalised to coprostanol) to calculate titres. As described earlier, use of an IS can account for variation introduced after its addition to the sample, facilitating comparison between different samples. Dai *et al.* (2014), Zhang *et al.* (2015), and Zhu *et al.* (2018) did not used an IS to quantify triterpenoid production, instead using only authentic standards. However, it was not clear if a standard curve was used or whether sample peak areas were compared to that of a single concentration of the standard.

While Kirby *et al.* (2008) used an IS, it was unclear if this was used as part of the quantification procedure, which involved the use of a standard curve to quantify β-amyrin production. The IS used was cholesterol rather than coprostanol, which is structurally similar to coprostanol and is also not produced by yeast, making it a suitable IS. Due to their similar structures, both cholesterol and coprostanol are likely to perform similarly as internal standards, although this was not tested in the present study. Several other studies also used cholesterol as an IS (Moses et al., 2014c; Arendt et al., 2017; Miettinen et al., 2017). Moses *et al.* (2014c) and Miettinen *et al.* (2017) normalised triterpenoids to the IS to allow comparison between different strains, but neither quantified production. Meanwhile, Arendt *et al.* (2017) used a standard curve of hederagenin to quantify medicagenic acid titres, however, it was unclear if the IS was used in these calculations. In addition, the use of a hederagenin standard to quantifiy medicagenic acid production may cause inaccuracies; it was observed above (Figure 3.5 A) that different triterpenoids can give different peak

areas when run at the same concentration. Thus, the standard used for quantification should ideally match the metabolite being quantified.

Chapter 4: Comparison of β-amyrin Synthases

4.1 Introduction

Beta-amyrin is one of the most widespread triterpenes and is produced by many plants (Hoshino, 2017; Vincken et al., 2007). Numerous enzymes that act on β-amyrin and its derivatives have been identified, and the biosynthesis of β-amyrin-derived triterpenoids (oleanane triterpenoids) is relatively well characterised. The oleanane triterpenoids contain numerous bioactive molecules that are of commercial and medical interest. These include triterpenoids that act as feeding deterrents towards plant pests (e.g. hederagenin cellobioside) (Nielsen et al., 2010; Augustin et al., 2012), serve as vaccine adjuvants (QS-21) (Ragupathi et al., 2011; Didierlaurent et al., 2017) or have anti-cancer activities (hederacolchiside A) (Bang et al., 2005). Indeed, oleanane triterpenoids and their derivatives have been evaluated in clinical trials for leukaemia, chronic kidney disease, Friedreich's ataxia, and in a vaccine against malaria (Tsao et al., 2010; Moses et al., 2013; Strawser et al., 2017; Gosling and von Seidlein, 2016; Mahmoudi and Keshavarz, 2017). For more information, see section 1.2 and Figure 1.6.

Although oleanane triterpenoids are of great industrial and medical interest, their commercial exploitation is currently hindered due to the difficulty in obtaining them in sufficiently high quantities. Plants typically produce low quantities of triterpenoids and are often not amenable to large-scale cultivation, making this method of production uneconomical. Furthermore, although chemical synthesis has proved effective for the production of a range of molecules, the complex structures of triterpenoids precludes their effective chemical synthesis. By contrast, the microbial production of triterpenoids, in which the plant biosynthetic pathways are expressed in a selected microorganism, has potential for the industrial production of triterpenoids. The yeast *Saccharomyces cerevisiae* is a particularly promising host, as it naturally produces the triterpene precursor 2,3-oxidosqualene, can functionally express plant cytochromes P450, and has been successfully employed to produce triterpenoids (Moses et al., 2014b; Fukushima et al., 2013; Arendt et al., 2017; Miettinen et al., 2017).

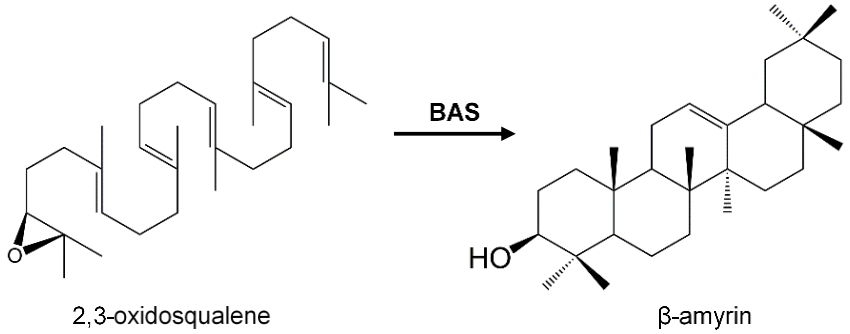


Figure 4.1. Cyclisation of 2,3-oxidosqualene to β-amyrin by a β-amyrin synthase (BAS).

The relatively well characterised pathways and the numerous bioactive molecules makes the oleananes excellent candidate triterpenoids for production in yeast. The monofunctional OSC responsible for β-amyrin biosynthesis (Figure 4.1) is β-amyrin synthase (BAS), many homologues of which have been identified from diverse plant species (Thimmappa et al., 2014); several multifunctional OSCs are also known to produce β-amyrin (Ghosh, 2016). When engineering yeast to produce oleanane triterpenoids, no studies have systemically compared the productivities of BASs when expressed in yeast. Instead, an arbitrary BAS variant is usually chosen (Fukushima et al., 2013; Kirby et al., 2008; Madsen, 2011; Seki et al., 2011; Miettinen et al., 2017). However, the choice of homologue can have a large impact on productivity. Indeed, Moses *et al.* compared GgBAS (from *Glycyrrhiza glabra*) and MtBAS (from *Medicago truncatula*) in yeast and obtained ~ 2-fold higher β-amyrin titres from the strain expressing GgBAS (Moses et al., 2014b). Similarly, Dai *et al.* observed ~ 50-fold higher titres for GgBAS compared with PgBAS1 (from *Panax ginseng*) when expressed in yeast (Dai et al., 2014).

To optimise this critical step in triterpenoid biosynthesis, twelve BAS homologues (Table 4.1) were compared for β-amyrin production in the haploid *S. cerevisiae* strain BY4741. The chosen BASs represent those most commonly used in metabolic engineering studies as well as several others. These included MtBAS, GgBAS, PgBAS1 and AaBAS, for which β-amyrin titres have previously been reported in yeast expressing these BASs. Moses *et al.* reported β-amyrin titres of 19 mg/L and 36 mg/L for MtBAS and GgBAS, respectively (Moses et al., 2014b), while Dai *et al*. reported β-amyrin titres of 107 mg/L when using GgBAS (Dai et al., 2014). PgBAS1 showed lower productivity, producing only 1.9 mg/L in the system used by Dai *et al.*, 2014.

AaBAS was used in the first reported production of β-amyrin in yeast, with titres of 6 mg/L being reported (Kirby et al., 2008). Notably, these yeast strains were metabolically engineered to enhance triterpenoid production, so the titres achieved by expressing a BAS alone may be lower. In addition to these BASs, seven other characterised BASs for which β-amyrin titres had not been reported (AsBAS, BvBAS, CqBAS1, EtBAS, LjBAS, PtBAS and SlBAS) were also chosen to ensure good coverage across species. Finally, an additional BAS was newly identified from *Gossypium hirsutum* (upland cotton) and functionally characterised.

Beta-amyrin production was quantified using GC-MS, and growth curves were recorded using a microplate assay to determine if BAS expression caused any detrimental effects on cell growth rate. The enzyme sequences were then analysed to investigate if the observed differences in activity could be explained by amino acid variants.

4.2 Discovery and characterisation of a specific β-amyrin synthase from *Gossypium hirsutum*

An initial investigation was undertaken to identify additional monofunctional β-amyrin synthase (BAS) homologues. A BLASTp search was performed on NCBI with default parameters against the non-redundant database, using a previously characterised monofunctional BAS from *Glycyrrhiza glabra* (GgBAS) as the query sequence. This returned 100 hits with amino acid sequence identities to GgBAS > 80 %, the majority of which were annotated as predicted BASs, but some predicted multifunctional OSCs and other triterpene synthases were also present. Two predicted monofunctional BASs were selected for analysis: GhBAS from *Gossypium hirsutum* (upland cotton) (XP_016749748) and TcBAS from *Theobroma cacao* (cocoa tree) (XP_017979420). Both had high sequence identities to previously characterised BASs (> 79 % identity to AaBAS, GgBAS, MtBAS and EtBAS) and contained the expected conserved residues for a monofunctional BAS (Figure 4.2) (Hoshino, 2017). These included the highly conserved DCTAE motif involved in initiation of the cyclisation reaction (Ito et al., 2013; Hoshino, 2017), the MWCYCR motif important for product specificity (Kushiro et al., 2000), and the QW motifs that are characteristic of OSCs and may be important for the stability of the protein (Poralla, 1994; Wendt et al., 1997; Salmon et al., 2016).

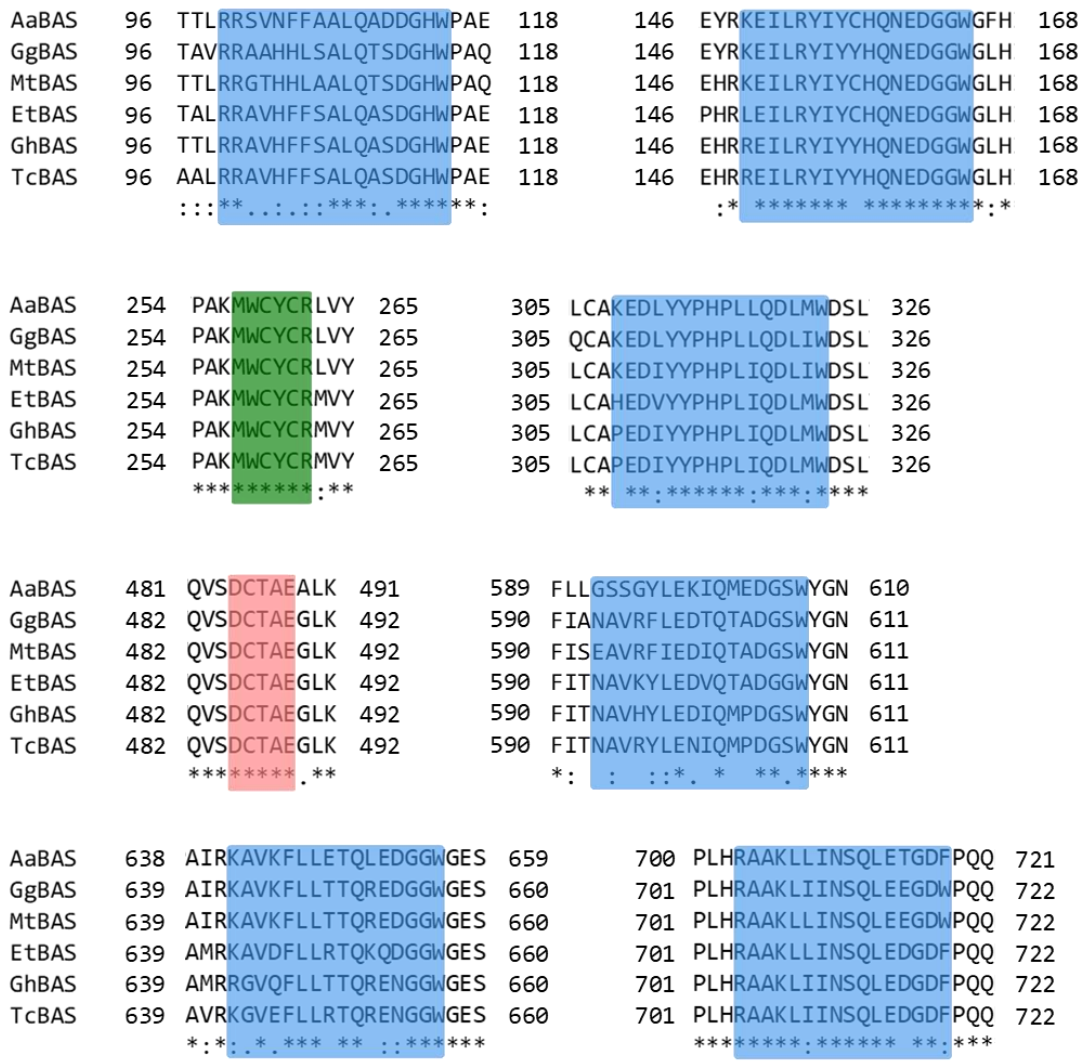


Figure 4.2. Multiple sequence alignment of GhBAS and TcBAS showing selected conserved residues. GhBAS and TcBAS were aligned with the previously characterised β-amyrin synthases AaBAS, GgBAS, MtBAS and EtBAS to determine if they contained the conserved residues expected for a β-amyrin synthase. The QW motifs (blue), MWCYCR motif (green) and DCTAE motif (red) are highlighted. The concensus is shown under the alignment: identical residues are marked by '*', conservative substitutions by '.', and semi-conservative substitutions by ':'. Protein sequences were aligned using Clustal Omega (www.ebi.ac.uk/Tools/msa/clustalo/) with default parameters. AaBAS: *Artemisia annua* β-amyrin synthase (BAS) (ACA13386); GgBAS: *Glycyrrhiza glabra* BAS (BAA89815); MtBAS: *Medicago truncatula* BAS (CAD23247); EtBAS: *Euphorbia tirucalli* BAS (BAE43642); GhBAS: putative *Gossypium hirsutum* BAS (XP_016749748); TcBAS: putative *Theobroma cacao* BAS (XP_017979420).

Triterpenoid biosynthetic pathways have not been reported for either *G. hirsutum* or *T. cacao*. However, β-amyrin has been reported in *Gossypium* spp. (Shakhidoyatov et al., 1997), suggesting that *G. hirsutum* may encode a BAS (although it is also possible that it instead encodes a multifunctional triterpene synthase that produces multiple triterpenes including β-amyrin). While *T. cacao* has been reported to contain triterpenoids and saponins, their identities were not determined (Falade et al., 2005; Zainal et al., 2014; Komlaga and Cojean, 2015). Nevertheless, due to the widespread occurrence of β-amyrin, it is likely that *T. cacao* produces this triterpene.

The *GhBAS* and *TcBAS* genes were codon-optimised for *S. cerevisiae*, synthesised, and assembled into a low copy yeast plasmid (CEN6/ARS4) for expression from the *GAL1* galactose-inducible promoter (*pGAL1*) (see materials and methods). Unfortunately, *TcBAS* could not be cloned into the MoClo-YTK entry vector; all obtained clones contained missense mutations or deletions and *TcBAS* was therefore not analysed. The *GhBAS* expression plasmid was transformed into *S. cerevisiae* BY4741 (strain MD-BA9; see Chapter 2, List of Yeast Strains) and analysed for β-amyrin production by thin layer chromatography (TLC). Cells were grown in SC-URA + 2 % D-Glc media for 24 h, then diluted into SC-URA + 2 % D-Gal media to induce gene expression and cultured for 96 h. Metabolites were extracted twice with hexane and loaded onto a TLC plate for analysis. An authentic β-amyrin standard and an extract from a strain carrying an empty vector (MD-N1) were also run as controls. The metabolite extraction and TLC methods were based on protocols provided by Thomas Louveau (Anne Osbourn lab, John Innes Centre). Several distinct bands were observed for MD-BA9, one of which had a similar retention time to the band observed for the β-amyrin standard (Figure 4.3). While a band with a similar retention time was observed for the control strain, it was much fainter compared with MD-BA9 and the intensities of the other bands were similar for both strains, indicating that the amount of total metabolite loaded was similar for each strain. Thus, the faint band for MD-N1 may correspond to a native yeast metabolite. The TLC results therefore indicate that MD-BA9 produces β-amyrin.

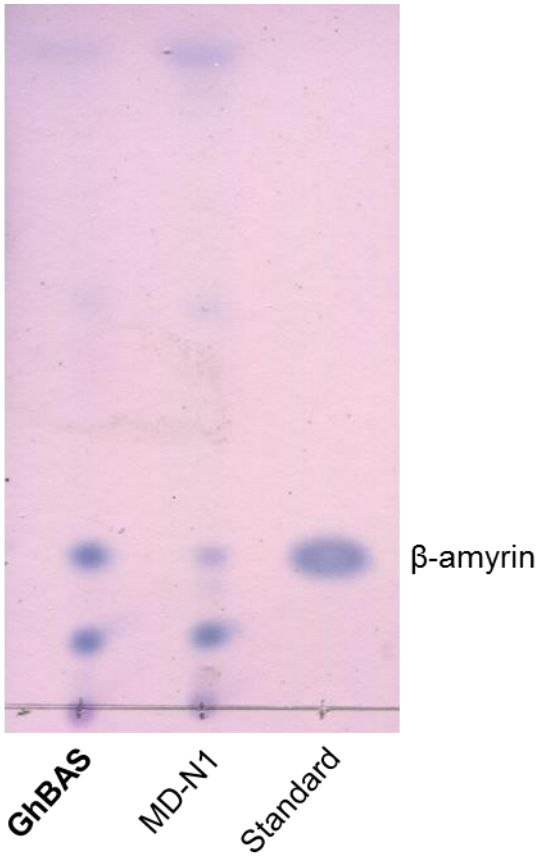


Figure 4.3. Thin layer chromatography (TLC) suggests that GhBAS produces β-amyrin. Thin layer chromatography was performed on extracts of MD-BA9 (expressing *GhBAS*) and MD-N1 (empty vector control) and on 0.6 µg β-amyrin standard. A band of the expected size for β-amyrin was observed for the strain expressing *GhBAS*, while MD-N1 had a fainter band of this size, suggesting that GhBAS may produce β-amyrin. Control: MD-N1; Standard: authentic β-amyrin standard.

To confirm β-amyrin production, an extract of MD-BA9 was analysed by GC-MS. By comparing the retention time and mass spectrum of the putative β-amyrin peak with an authentic standard, this should further verify the presence of β-amyrin in the sample. As a positive control, extract from a strain expressing a previously verified BAS (*GgBAS*; strain MD-BA3) was also analysed by GC-MS, together with the negative control strain MD-N1 that expressed no BAS. In addition, a mixture of authentic triterpenoid standards was analysed. The standards mixture was as described before, containing coprostanol, ergosterol, β-amyrin, erythrodiol and oleanolic acid (see Chapter 3, section 3.4). Analysis of the total ion chromatograms (TICs) showed that MD-BA9 had distinct peaks that corresponded in retention time to ergosterol and β-amyrin from the standards mix (Figure 4.4 A). As expected, these peaks were also present for MD-BA3, while only the ergosterol peak was observed for MD-N1. The putative β-amyrin peak for MD-BA9 was confirmed to be β-amyrin by comparison of its mass spectrum with the standard (Figure 4.4 B). As expected, the characteristic fragment ions derived from the retro Diels-Alder fragmentation of the parent ion (with m/z values of 218, 203 and 189) were present in high abundance, while the parent ion and those fragment ions deriving from the loss of TMSOH and/or a methyl group from it (with m/z values of 498, 483, 408 and 393) were in low abundance.

All three strains had an additional peak at ~ 21.5 min, as well as some minor peaks that were not present in the standards mix (Figure 4.4 A). These likely correspond to native yeast metabolites. As expected, no erythrodiol or oleanolic acid were observed from the yeast strain extracts. While MD-N1 showed a small peak with a similar retention time as the oleanolic acid standard, its mass spectrum was not consistent with this triterpenoid. Because this peak was not observed for MD-BA9 or MD-BA3, it may derive from carryover of the aqueous phase or emulsion during metabolite extraction, which would contain metabolites not present in the organic phase, or from uneven extraction or loading across the yeast samples. In summary, the GC-MS results confirm that GhBAS is a genuine β-amyrin synthase that specifically produces β-amyrin when expressed in yeast.

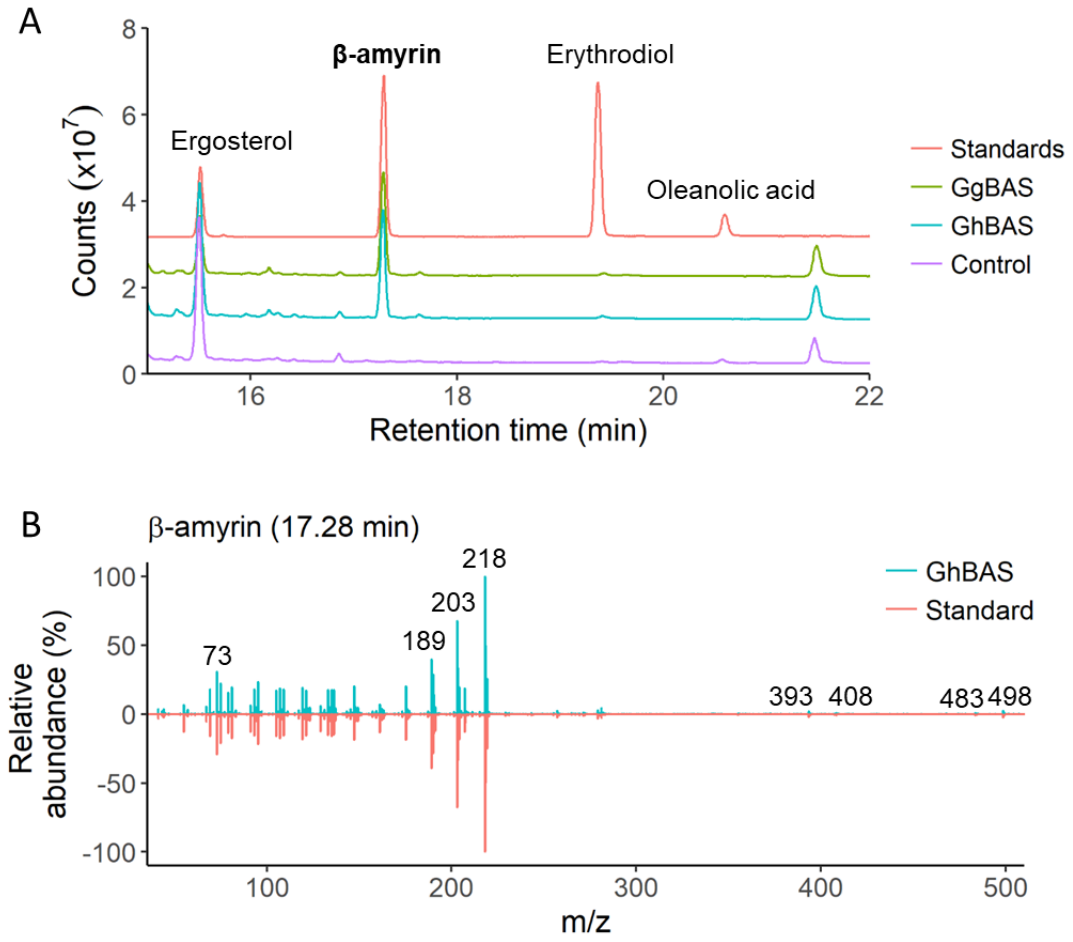


Figure 4.4. GC-MS confirming β-amyrin production by GhBAS. (A) Total ion chromatograms (TICs) zoomed between 15-22 min from GC-MS analysis of MD-BA9 (*GhBAS*), MD-BA3 (*GgBAS*), MD-N1 (empty vector control) and a mixture of triterpenoid standards containing β-amyrin, erythrodiol and oleanolic acid at 50 µg/ml concentration. MD-BA3 is a positive control that expresses *GgBAS*, a previously characterised β-amyrin synthase. **(B)** Mass spectra of the β-amyrin peaks for *GhBAS* and the β-amyrin standard, with the m/z values of key ions labelled. Analysis of the TICs and mass spectra confirmed that GhBAS produces β-amyrin. Emily Johnston set up the galactose cultures and Tessa Moses performed the extractions.

4.3 Comparison of β-amyrin synthase homologues for β-amyrin production

A total of twelve BASs were selected for comparison in yeast (Table 4.1). These included those previously used in metabolic engineering and yeast expression studies (e.g. AaBAS, GgBAS, MtBAS and PgBAS1) (Kirby et al., 2008; Moses et al., 2014b, 2015b; Dai et al., 2014; Arendt et al., 2017), the newly discovered GhBAS, and seven others to ensure good coverage across species. A protein sequence alignment of the BASs and other OSCs is shown in Figure 4.5. All BASs shared ≥ 70 % amino acid sequence identity except for AsBAS, which had ~ 50 % sequence identity to the other BASs (Figure 4.6). Phylogenetic analysis indicated that AsBAS was less closely related to the other BASs (Figure 4.7), which is consistent with previous phylogenetic analyses (Thimmappa et al., 2014). All of the chosen BASs except AsBAS are from dicotyledonous plants, with AsBAS being from the monocotyledon *Avena strigosa*. AsBAS is the only characterised monocot BAS, and is believed to have evolved separately from the dicot BASs (Thimmappa et al., 2014; Haralampidis et al., 2001; Qi et al., 2004).

Table 4.1. β-amyrin synthase homologues tested in this study

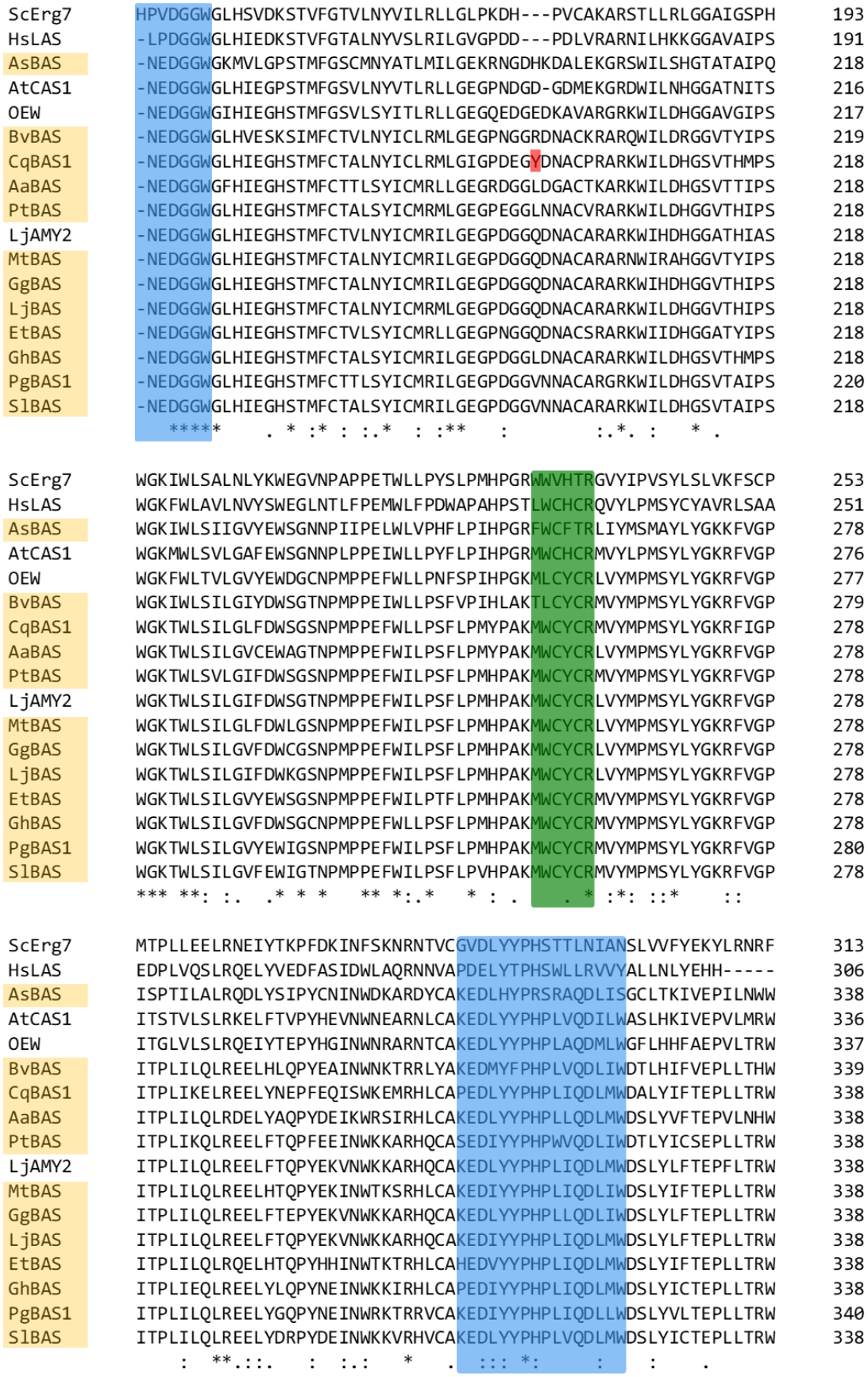
BAS	Species	Accession number*	Reference
AsBAS	*Avena strigosa*	CAC84558.1	Haralampidis *et al.*, 2001
PgBAS1	*Panax ginseng*	BAA33461.1	Kushiro, Shibuya and Ebizuka, 1998
GgBAS	*Glycyrrhiza glabra*	BAA89815.1	Hayashi *et al.*, 2001
LjBAS	*Lotus japonicus*	BAE53429.1	Iturbe-Ormaetxe *et al.*, 2003
EtBAS	*Euphorbium tirucalli*	BAE43642.1	Kajikawa *et al.*, 2005
PtBAS	*Polygala tenuifolia*	ABL07607.1	Jin *et al.*, 2014
GhBAS	*Gossypium hirsutum*	XP_016749748.1	The present study
AaBAS	*Artemisia annua*	ACA13386.1	Kirby *et al.*, 2008
BvBAS	*Barbarea vulgaris*	AFF27505.1	Wei *et al.*, 2012
CqBAS1	*Chenopodium quinoa*	ANY30852.1	Fiallos-Jurado *et al.*, 2016
MtBAS	*Medicago truncatula*	CAD23247.1	Suzuki *et al.*, 2002
SlBAS	*Solanum lycopersicum*	ADU52574.1	Wang *et al.*, 2011

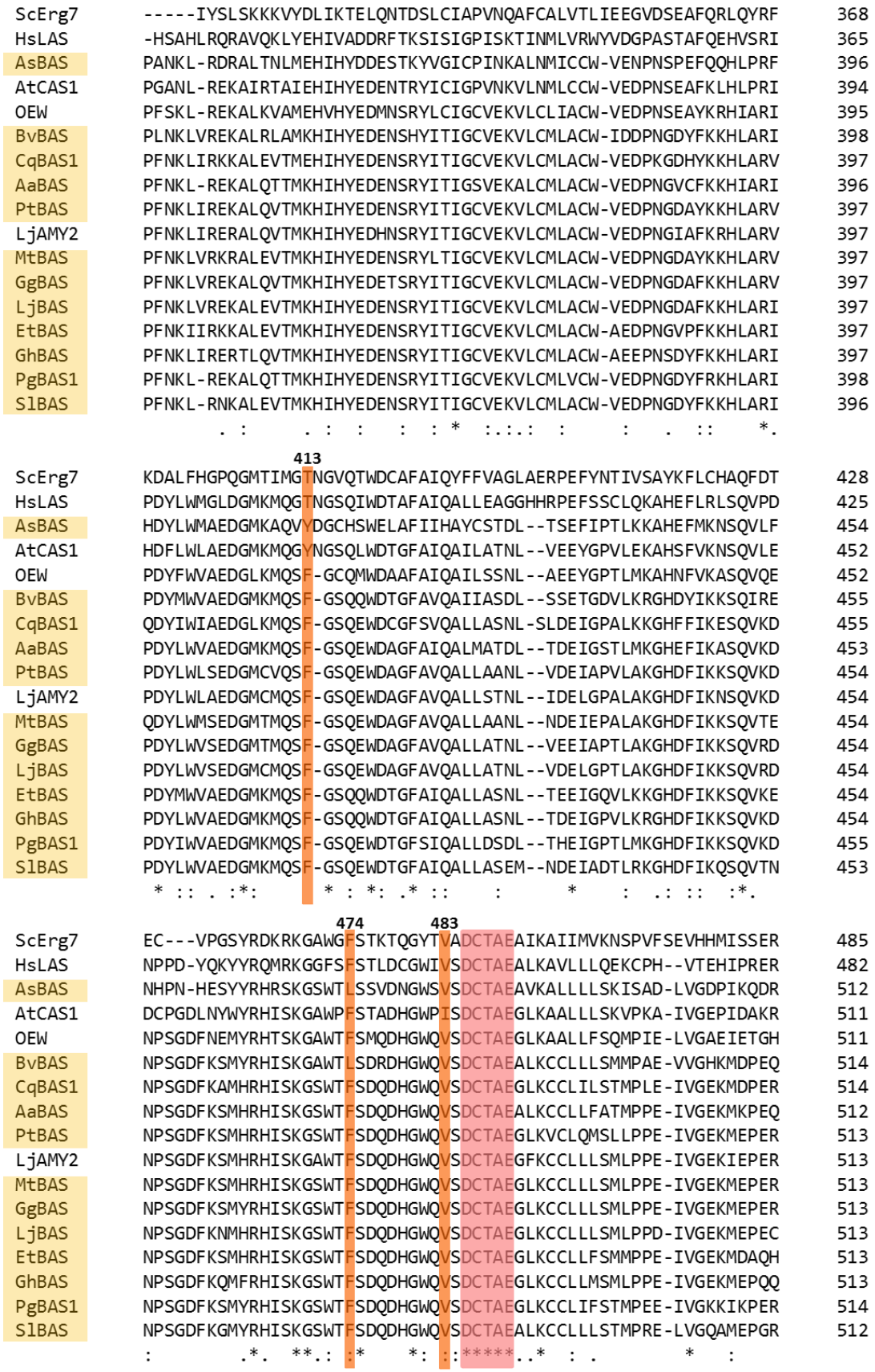
* Note that AaBAS, BvBAS, CqBAS1, MtBAS and SlBAS had an additional 'GS' sequence at the C-terminus, and CqBAS1 also contained a D196Y mutation. Sequences were otherwise identical to those from GenBank.

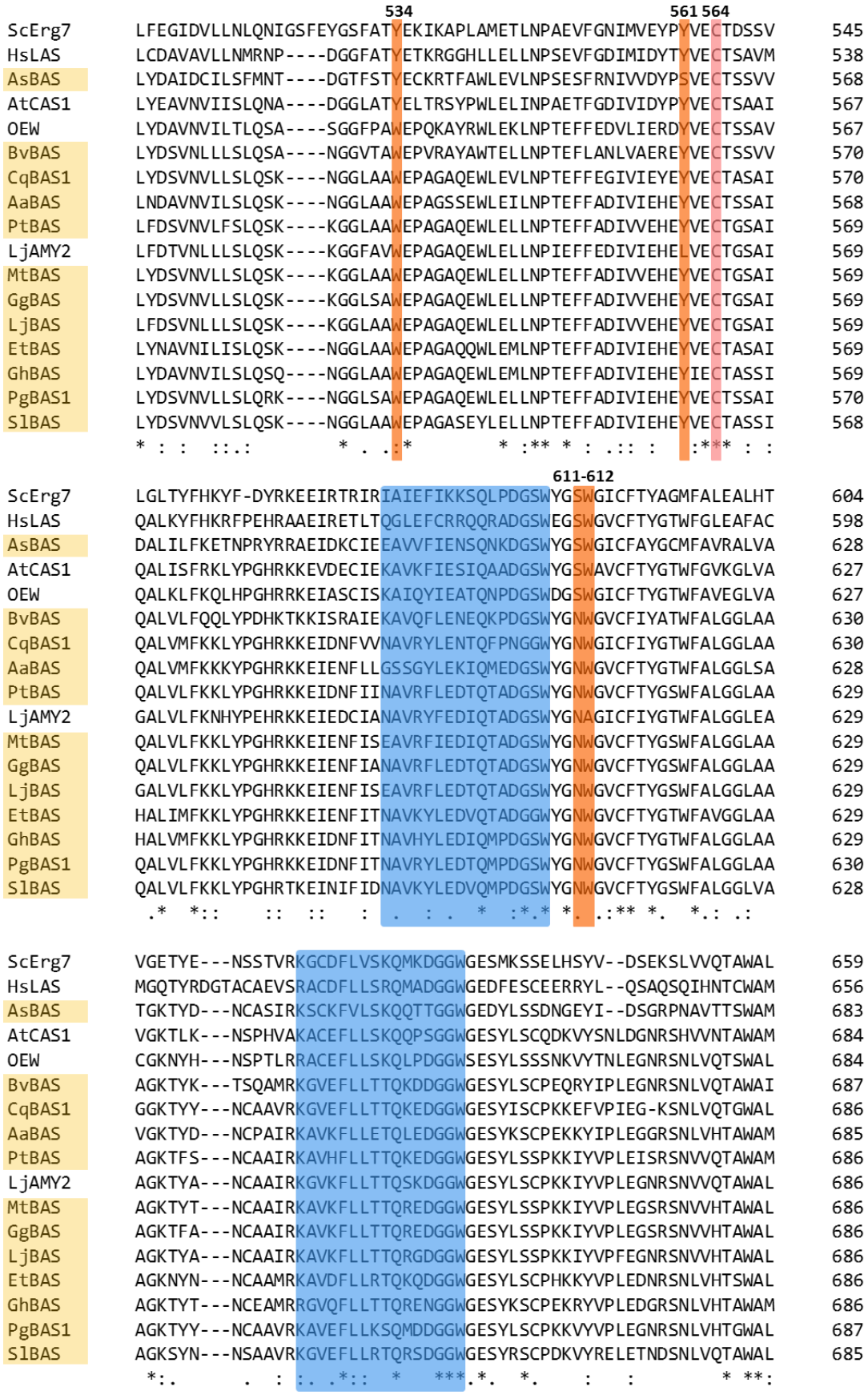
```
ScErg7      MTEFYSDTIGLP-----KTDPRL---------WRLRTDELGRESWEYLTPQQAANDP---      43
HsLAS       --------MTEGTCLRRRGGPYKTEPATDLGRW-RLNCERGRQTWTYLQDERA------G      45
AsBAS       ---MWRLTIGEGG------GPWL---------K-SNNGFLGRQVWEYDADAGT-PEERAE      40
AtCAS1      ---MWKLKIAEGG------SPWL---------R-TTNNHVGRQFWEFDPNLGT-PEDLAA      40
OEW         ---MWKLKIADGT------GPWL---------T-TTNNHIGRQHWEFDPEAGT-PDERVE      40
BvBAS       ---MWRLKLGEGN----GDDPYL---------F-SSNNFVGRQTWEFDPKAGT-LEERAA      42
CqBAS1      ---MWRLKVGEGA-----NDPYL---------Y-STNNFVGRQTWEFDPNYGT-PEEREE      41
AaBAS       ---MWRLKIAEGR-----NDPYL---------Y-STNNFVGRQIWEFDPNYGT-PEERAE      41
PtBAS       ---MWRLKVGEGK-----NDPYL---------F-STNDYTGRQTWEFDPDAGT-PEERAE      41
LjAMY2      ---MWKLKVADGG-----KNPYI---------F-SINNFVGRQTWEYDPDAGT-PEERAQ      41
MtBAS       ---MWKLKIGEGK-----NEPYL---------F-STNNFVGRQTWEYDPEAGS-EEERAQ      41
GgBAS       ---MWRLKIAEGG-----KDPYI---------Y-STNNFVGRQTWEYDPDGGT-PEERAQ      41
LjBAS       ---MWKLKVADGG-----KDPYI---------F-STNNFVGRQTWEYDPDAGT-PEERAQ      41
EtBAS       ---MWKLKIAEGG-----NDEYL---------Y-STNNYVGRQTWVFDPQPPT-PQELAQ      41
GhBAS       ---MWKLKVAEGV-----DGPYL---------Y-STNNYVGRQTWEFDPDSGT-PEERAQ      41
PgBAS1      ---MWKLKIAEGN----KNDPYL---------Y-STNNFVGRQTWEFDPDYVASPGELEE      43
SlBAS       ---MWKLKIAEGQ-----NGPYL---------Y-STNNYVGRQTWEFDPNGGTI-EERAK      41
                       :                           .   **: * :     :

ScErg7      ---------------PSTFTQW---LLQDPKFPQPHPER-----NKHSPDFSAFDACHN      79
HsLAS       REQTGLEAYA-------------LGL-DTKNYFKDLPK----------AHTAFEGALN      79
AsBAS       VERVRAEFTKNRFQRKESQDLLLRLQYAKDNPLPANIPTEAKLEKSTEVTHETIYESLMR      100
AtCAS1      VEEARKSFSDNRFVQKHSADLLMRLQFSRENLISPVLPQ-VKIEDTDDVTEEMVETTLKR      99
OEW         VERLREEFKKNRFRTKQSADLLMRMQLVKENQRVQIPPA-IKIKETEGITEEAVITTLRR      99
BvBAS       VEEARRSFLVNRSRVKACSDLLWRMQFLKEAKFEQVIPP-VKIEDAKDITYENATDSLRR      101
CqBAS1      VEEARRNFYNNRFKVKPCGDLIWRLQFLREKNFKQTIPQ-VKVEEGEEITYETATTTLKR      100
AaBAS       VEQARVDFWNHRHEVKPSSDVLWRMQFLREKGFEQTIPQ-VKIEDGEEISYEKATTTLRR      100
PtBAS       VEAARQAFYDNRFQFKNCGDLLWRFQFLRDKNFKQTIPK-VKVEDGQQITYEMATDTVRR      100
LjAMY2      VEEARQDFYNNRYKVKTCGDRLWRFQVMRENNFKQTIPS-VKIEDGEKVTYDKVTTTVRR      100
MtBAS       VEEARKNFYDNRFKVKPCGDLLWRFQVLRENNFMQTIDG-VKIEDGEEITYEKATTTLRR      100
GgBAS       VDAARLHFYNNRFQVKPCGDLLWRFQILRENNFKQTIAS-VKIGDGEEITYEKATTAVRR      100
LjBAS       VEEARQDFYNNRYKVKPCGDLLWRFQVLRENNFKQTIPS-VKIEDGEEITYEKATTTLKR      100
EtBAS       VQQARLNFYNNRYHVKPSSDLLWRFQFLREKNFKQTIPQ-AKINEGEDITYEKATTALRR      100
GhBAS       VEEARKNFYKNRHHVKPSADLLWRMQFLKEKNFKQSIPA-VKIDDGEQITYEKATTTLRR      100
PgBAS1      VEQVRRQFWDNRYQVKPSGDLLWRMQFLREKNFRQTIPQ-VKVGDDEAVTYEAATTTLRR      102
SlBAS       IEEARQQFWNNRYKVKPSSDLLWRIQFLGEKNFKQKIPA-VKVEEGEEISHEVATIALHR      100
                                                                     .

ScErg7      GASFFKLLQEPDSGIFPCQYKGPMFMTIGYVAVNYIAG---IEIPEHERIELIRYIVNTA      136
HsLAS       GMTFYVGLQA-EDGHWTGDYGGPLFLLPGLLITCHVA---RIPLPAGYREEIVRYLRSVQ      135
AsBAS       ALHQYSSLQA-DDGHWPGDYSGILFIMPIIIFSLYVTRSLDTFLSPEHRHEICRYIYNQQ      159
AtCAS1      GLDFYSTIQA-HDGHWPGDYGGPMFLLPGLIITLSITGALNTVLSEQHKQEMRRYLYNHQ      158
OEW         AISFYSTIQA-HDGHWPAESAGPLFFLPPLVLALYVTGAINVVLSREHQKEITRYIYNHQ      158
BvBAS       GVSFFSALQA-SDGHWPGEIAGPLFFLPPLVFCLYITGHLEEIFDEEHRKEMLRHVYCHQ      160
CqBAS1      AVNVFTALQS-DQGHWPAEIAGPQFFLPPLVFCLYITGDLNSVFGPEHRREILRSIYYHQ      159
AaBAS       SVNFFAALQA-DDGHWPAENAGPLYFMQPLVICLYITGHLNTVFPAEYRKEILRYIYCHQ      159
PtBAS       AAHHLGGLQS-SHGHWPAQIAGPLFFMPPLVFCLYITGHLNTVFPEEHRKEILRYIYYHQ      159
LjAMY2      AAHHLAGLQT-SDGHWPAQIAGPLLFTPPLIFCMYITGHLDSVFPEVYRKEILRYTYVHQ      159
MtBAS       GTHHLAALQT-SDGHWPAQIAGPLFFMPPLVFCVYITGHLDSVFPREHRKEILRYIYCHQ      159
GgBAS       AAHHLSALQT-SDGHWPAQIAGPLFFLPPLVFCMYITGHLDSVFPEEYRKEILRYIYYHQ      159
LjBAS       AAHHLAALQT-SDGHWPAQIAGPLFFQPPLVFCMYITGHLNSVFPEEYRKEILRYIYVHQ      159
EtBAS       AVHFFSALQA-SDGHWPAENAGPLFFLPPLVMCLYITGHLDTVFPAPHRLEILRYIYCHQ      159
GhBAS       AVHFFSALQA-SDGHWPAENAGPLFFLPPLVFSMYITGHLNTVFPEEHRREILRYIYYHQ      159
PgBAS1      AVHFFSALQA-SDGHWPAENSGPLFFLPPLVMCVYITGHLDTVFPAEHRKEILRYIYCHQ      161
SlBAS       AVNFFSALQA-TDGHWPAENAGPLFFLPPLVMCMYITGHLNTVFPAEHRKEILRYIYCHQ      159
            .      :*    * :  :  *   :    :    ::      :    : *: *
```


ScErg7 HPVDGGWGLHSVDKSTVFGTVLNYVILRLLGLPKDH---PVCAKARSTLLRLGGAIGSPH 193
HsLAS -LPDGGWGLHIEDKSTVFGTALNYVSLRILGVGPDD---PDLVRARNILHKKGGAVAIPS 191
AsBAS -NEDGGWGKMVLGPSTMFGSCMNYATLMILGEKRNGDHKDALEKGRSWILSHGTATAIPQ 218
AtCAS1 -NEDGGWGLHIEGPSTMFGSVLNYVTLRLLGEGPNDGD-GDMEKGRDWILNHGGATNITS 216
OEW -NEDGGWGIHIEGHSTMFGSVLSYITLRLLGEGQEDGEDKAVARGRKWILDHGGAVGIPS 217
BvBAS -NEDGGWGLHVESKSIMFCTVLNYICLRMLGEGPNGGRDNACKRARQWILDRGGVTYIPS 219
CqBAS1 -NEDGGWGLHIEGHSTMFCTALNYICLRMLGIGPDEGYDNACPRARKWILDHGSVTHMPS 218
AaBAS -NEDGGWGFHIEGHSTMFCTTLSYICMRLLGEGRDGGLDGACTKARKWILDHGSVTTIPS 218
PtBAS -NEDGGWGLHIEGHSTMFCTALSYICMRMLGEGPEGGLNNACVRARKWILDHGGVTHIPS 218
LjAMY2 -NEDGGWGLHIEGHSTMFCTVLNYICMRILGEGPDGGQDNACARARKWIHDHGGATHIAS 218
MtBAS -NEDGGWGLHIEGHSTMFCTALNYICMRILGEGPDGGQDNACARARNWIRAHGGVTYIPS 218
GgBAS -NEDGGWGLHIEGHSTMFCTALNYICMRILGEGPDGGQDNACARARKWIHDHGGVTHIPS 218
LjBAS -NEDGGWGLHIEGHSTMFCTALNYICMRMLGEGPDGGQDNACARARKWILDHGGVTHIPS 218
EtBAS -NEDGGWGLHIEGHSTMFCTVLSYICMRLLGEGPNGGQDNACSRARKWIIDHGGATYIPS 218
GhBAS -NEDGGWGLHIEGHSTMFCTALSYICMRILGEGPDGGLDNACARARKWILDHGSVTHMPS 218
PgBAS1 -NEDGGWGLHIEGHSTMFCTTLSYICMRILGEGPDGGVNNACARGRKWILDHGSVTAIPS 220
SlBAS -NEDGGWGLHIEGHSTMFCTALSYICMRILGEGPDGGVNNACARARKWILDHGSVTAIPS 218
WGKIWLSALNLYKWEGVNPAPPETWLLPYSLPMHPGRWWVHTRGVYIPVSYLSLVKFSCP 253
WGKFWLAVLNVYSWEGLNTLFPEMWLFPDWAPAHPSTLWCHCRQVYLPMSYCYAVRLSAA 251
WGKIWLSIIGVYEWSGNNPIIPELWLVPHFLPIHPGRFWCFTRLIYMSMAYLYGKKFVGP 278
WGKMWLSVLGAFEWSGNNPLPPEIWLLPYFLPIHPGRMWCHCRMVYLPMSYLYGKRFVGP 276
WGKFWLTVLGVYEWDGCNPMPPEFWLLPNFSPIHPGKMLCYCRLVYMPMSYLYGKRFVGP 277
WGKIWLSILGIYDWSGTNPMPPEIWLLPSFVPIHLAKTLCYCRMVYMPMSYLYGKRFVGP 279
WGKTWLSILGLFDWSGSNPMPPEFWLLPSFLPMYPAKMWCYCRMVYMPMSYLYGKRFIGP 278
WGKTWLSILGVCEWAGTNPMPPEFWILPSFLPMYPAKMWCYCRLVYMPMSYLYGKRFVGP 278
WGKTWLSVLGIFDWSGSNPMPPEFWILPSFLPMHPAKMWCYCRMVYMPMSYLYGKRFVGP 278
WGKTWLSILGIFDWSGTNPMPPEFWILPSFLPMHPAKMWCYCRLVYMPMSYLYGKRFVGP 278
WGKTWLSILGLFDWLGSNPMPPEFWILPSFLPMHPAKMWCYCRLVYMPMSYLYGKRFVGP 278
WGKTWLSILGVFDWCGSNPMPPEFWILPSFLPMHPAKMWCYCRLVYMPMSYLYGKRFVGP 278
WGKTWLSILGIFDWKGSNPMPPEFWILPSFLPMHPAKMWCYCRLVYMPMSYLYGKRFVGP 278
WGKTWLSILGVYEWSGSNPMPPEFWILPTFLPMHPAKMWCYCRMVYMPMSYLYGKRFVGP 278
WGKTWLSILGVFDWSGCNPMPPEFWLLPSFLPMHPAKMWCYCRMVYMPMSYLYGKRFVGP 278
WGKTWLSILGVYEWIGSNPMPPEFWILPSFLPMHPAKMWCYCRMVYMPMSYLYGKRFVGP 280
WGKTWLSILGVFEWIGTNPMPPEFWILPSFLPVHPAKMWCYCRMVYMPMSYLYGKRFVGP 278
MTPLLEELRNEIYTKPFDKINFSKNRNTVCGVDLYYPHSTTLNIANSLVVFYEKYLRNRF 313
EDPLVQSLRQELYVEDFASIDWLAQRNNVAPDELYTPHSWLLRVVYALLNLYEHH----- 306
ISPTILALRQDLYSIPYCNINWDKARDYCAKEDLHYPRSRAQDLISGCLTKIVEPILNWW 338
ITSTVLSLRKELFTVPYHEVNWNEARNLCAKEDLYYPHPLVQDILWASLHKIVEPVLMRW 336
ITGLVLSLRQEIYTEPYHGINWNRARNTCAKEDLYYPHPLAQDMLWGFLHHFAEPVLTRW 337
ITPLILQLREELHLQPYEAINWNKTRRLYAKEDMYFPHPLVQDLIWDTLHIFVEPLLTHW 339
ITPLIKELREELYNEPFEQISWKEMRHLCAPEDLYYPHPLIQDLMWDALYIFTEPLLTRW 338
ITPLILQLRDELYAQPYDEIKWRSIRHLCAKEDLYYPHPLLQDLMWDSLYVFTEPVLNHW 338
ITPLIKQLREELFTQPFEEINWKKARHQCASEDIYYPHPWVQDLIWDTLYICSEPLLTRW 338
ITPLILQLREELFTQPYEKVNWKKARHQCAKEDLYYPHPLIQDLMWDSLYLFTEPFLTRW 338
ITPLILQLREELHTQPYEKINWTKSRHLCAKEDIYYPHPLIQDLIWDSLYIFTEPLLTRW 338
ITPLILQLREELFTEPYEKVNWKKARHQCAKEDLYYPHPLLQDLIWDSLYLFTEPLLTRW 338
ITPLILQLREELFTQPYEKVNWKKARHQCAKEDIYYPHPLIQDLMWDSLYLFTEPLLTRW 338
ITPLILQLRQELHTQPYHHINWTKTRHLCAHEDVYYPHPLIQDLMWDSLYIFTEPLLTRW 338
ITPLIEQLREELYLQPYNEINWKKIRHLCAPEDIYYPHPLIQDLMWDSLYICTEPLLTRW 338
ITPLILQLREELYGQPYNEINWRKTRRVCAKEDIYYPHPLIQDLLWDSLYVLTEPLLTRW 340
ITPLILQLREELYDRPYDEINWKKVRHVCAKEDLYYPHPLVQDLMWDSLYICTEPLLTRW 338


413
474
483


534
561 564
611-612
ScErg7 LFEGIDVLLNLQNIGSFEYGSFATYEKIKAPLAMETLNPAEVFGNIMVEYPYVECTDSSV 545
HsLAS LCDAVAVLLNMRNP----DGGFATYETKRGGHLLELLNPSEVFGDIMIDYTYVECTSAVM 538
AsBAS LYDAIDCILSFMNT----DGTFSTYECKRTFAWLEVLNPSESFRNIVVDYPSVECTSSVV 568
AtCAS1 LYEAVNVIISLQNA----DGGLATYELTRSYPWLELINPAETFGDIVIDYPYVECTSAAI 567
OEW LYDAVNVILTLQSA----SGGFPAWEPQKAYRWLEKLNPTEFFEDVLIERDYVECTSSAV 567
BvBAS LYDSVNLLLSLQSA----NGGVTAWEPVRAYAWTELLNPTEFLANLVAEREYVECTSSVV 570
CqBAS1 LYDSVNVLLSLQSK----NGGLAAWEPAGAQEWLEVLNPTEFFEGIVIEYEYVECTASAI 570
AaBAS LNDAVNVILSLQSK----NGGLAAWEPAGSSEWLEILNPTEFFADIVIEHEYVECTSSAI 568
PtBAS LFDSVNVLFSLQSK----KGGLAAWEPAGAQEWLELLNPTEFFADIVVEHEYVECTGSAI 569
LjAMY2 LFDTVNLLLSLQSK----KGGFAVWEPAGAQEWLELLNPIEFFEDIVIEHELVECTGSAI 569
MtBAS LYDSVNVLLSLQSK----KGGLAAWEPAGAQEWLELLNPTEFFADIVVEHEYVECTGSAI 569
GgBAS LYDSVNVLLSLQSK----KGGLSAWEPAGAQEWLELLNPTEFFADIVVEHEYVECTGSAI 569
LjBAS LFDSVNLLLSLQSK----KGGLAAWEPAGAQEWLELLNPTEFFADIVVEHEYVECTGSAI 569
EtBAS LYNAVNILISLQSK----NGGLAAWEPAGAQQWLEMLNPTEFFADIVIEHEYVECTASAI 569
GhBAS LYDAVNVILSLQSQ----NGGLAAWEPAGAQEWLEMLNPTEFFADIVIEHEYIECTASSI 569
PgBAS1 LYDSVNVLLSLQRK----NGGLSAWEPAGAQEWLELLNPTEFFADIVIEHEYVECTSSAI 570
SlBAS LYDSVNVVLSLQSK----NGGLAAWEPAGASEYLELLNPTEFFADIVIEHEYVECTASSI 568
ScErg7 LGLTYFHKYF-DYRKEEIRTRIRIAIEFIKKSQLPDGSWYGSWGICFTYAGMFALEALHT 604
HsLAS QALKYFHKRFPEHRAAEIRETLTQGLEFCRRQQRADGSWEGSWGVCFTYGTWFGLEAFAC 598
AsBAS DALILFKETNPRYRRAEIDKCIEEAVVFIENSQNKDGSWYGSWGICFAYGCMFAVRALVA 628
AtCAS1 QALISFRKLYPGHRKKEVDECIEKAVKFIESIQAADGSWYGSWAVCFTYGTWFGVKGLVA 627
OEW QALKLFKQLHPGHRRKEIASCISKAIQYIEATQNPDGSWDGSWGICFTYGTWFAVEGLVA 627
BvBAS QALVLFQQLYPDHKTKKISRAIEKAVQFLENEQKPDGSWYGNWGVCFIYATWFALGGLAA 630
CqBAS1 QALVMFKKLYPGHRKKEIDNFVVNAVRYLENTQFPNGGWYGNWGICFIYGTWFALGGLAA 630
AaBAS QALVMFKKKYPGHRKKEIENFLLGSSGYLEKIQMEDGSWYGNWGVCFTYGTWFALGGLSA 628
PtBAS QALVLFKKLYPGHRKKEIDNFIINAVRFLEDTQTADGSWYGNWGVCFTYGSWFALGGLAA 629
LjAMY2 GALVLFKNHYPEHRKKEIEDCIANAVRYFEDIQTADGSWYGNAGICFIYGTWFALGGLEA 629
MtBAS QALVLFKKLYPGHRKKEIENFISEAVRFIEDIQTADGSWYGNWGVCFTYGSWFALGGLAA 629
GgBAS QALVLFKKLYPGHRKKEIENFIANAVRFLEDTQTADGSWYGNWGVCFTYGSWFALGGLAA 629
LjBAS GALVLFKKLYPGHRKKEIENFISEAVRFLEDTQTADGSWYGNWGVCFTYGSWFALGGLAA 629
EtBAS HALIMFKKLYPGHRKKEIENFITNAVKYLEDVQTADGGWYGNWGVCFTYGTWFAVGGLAA 629
GhBAS HALVMFKKLYPGHRKKEIDNFITNAVHYLEDIQMPDGSWYGNWGVCFTYGTWFALGGLAA 629
PgBAS1 QALVLFKKLYPGHRKKEIDNFITNAVRYLEDTQMPDGSWYGNWGVCFTYGSWFALGGLAA 630
SlBAS QALVLFKKLYPGHRTKEINIFIDNAVKYLEDVQMPDGSWYGNWGVCFTYGSWFALGGLVA 628
ScErg7 VGETYE---NSSTVRKGCDFLVSKQMKDGGWGESMKSSELHSYV--DSEKSLVVQTAWAL 659
HsLAS MGQTYRDGTACAEVSRACDFLLSRQMADGGWGEDFESCEERRYL--QSAQSQIHNTCWAM 656
AsBAS TGKTYD---NCASIRKSCKFVLSKQQTTGGWGEDYLSSDNGEYI--DSGRPNAVTTSWAM 683
AtCAS1 VGKTLK---NSPHVAKACEFLLSKQQPSGGWGESYLSCQDKVYSNLDGNRSHVVNTAWAM 684
OEW CGKNYH---NSPTLRRACEFLLSKQLPDGGWSESYLSSSNKVYTNLEGNRSNLVQTSWAL 684
BvBAS AGKTYK---TSQAMRKGVEFLLTTQKDDGGWGESYLSCPEQRYIPLEGNRSNLVQTAWAI 687
CqBAS1 GGKTYY---NCAAVRKGVEFLLTTQKEDGGWGESYISCPKKEFVPIEG-KSNLVQTGWAL 686
AaBAS VGKTYD---NCPAIRKAVKFLLETQLEDGGWGESYKSCPEKKYIPLEGGRSNLVHTAWAM 685
PtBAS AGKTFS---NCAAIRKAVHFLLTTQKEDGGWGESYLSSPKKIYVPLEISRSNVVQTAWAM 686
LjAMY2 AGKTYA---NCAAIRKGVKFLLTTQSKDGGWGESYLSCPKKIYVPLEGNRSNVVQTAWAL 686
MtBAS AGKTYT---NCAAIRKAVKFLLTTQREDGGWGESYLSSPKKIYVPLEGSRSNVVHTAWAL 686
GgBAS AGKTFA---NCAAIRKAVKFLLTTQREDGGWGESYLSSPKKIYVPLEGSRSNVVHTAWAL 686
LjBAS AGKTYA---NCAAIRKAVKFLLTTQRGDGGWGESYLSSPKKIYVPFEGNRSNVVHTAWAL 686
EtBAS AGKNYN---NCAAMRKAVDFLLRTQKQDGGWGESYLSCPHKKYVPLEDNRSNLVHTSWAL 686
GhBAS AGKTYT---NCEAMRRGVQFLLTTQRENGGWGESYKSCPEKRYVPLEDGRSNLVHTAWAM 686
PgBAS1 AGKTYY---NCAAVRKAVEFLLKSQMDDGGWGESYLSCPKKVYVPLEGNRSNLVHTGWAL 687
SlBAS AGKSYN---NSAAVRKGVEFLLRTQRSDGGWGESYRSCPDKVYRELETNDSNLVQTAWAL 685

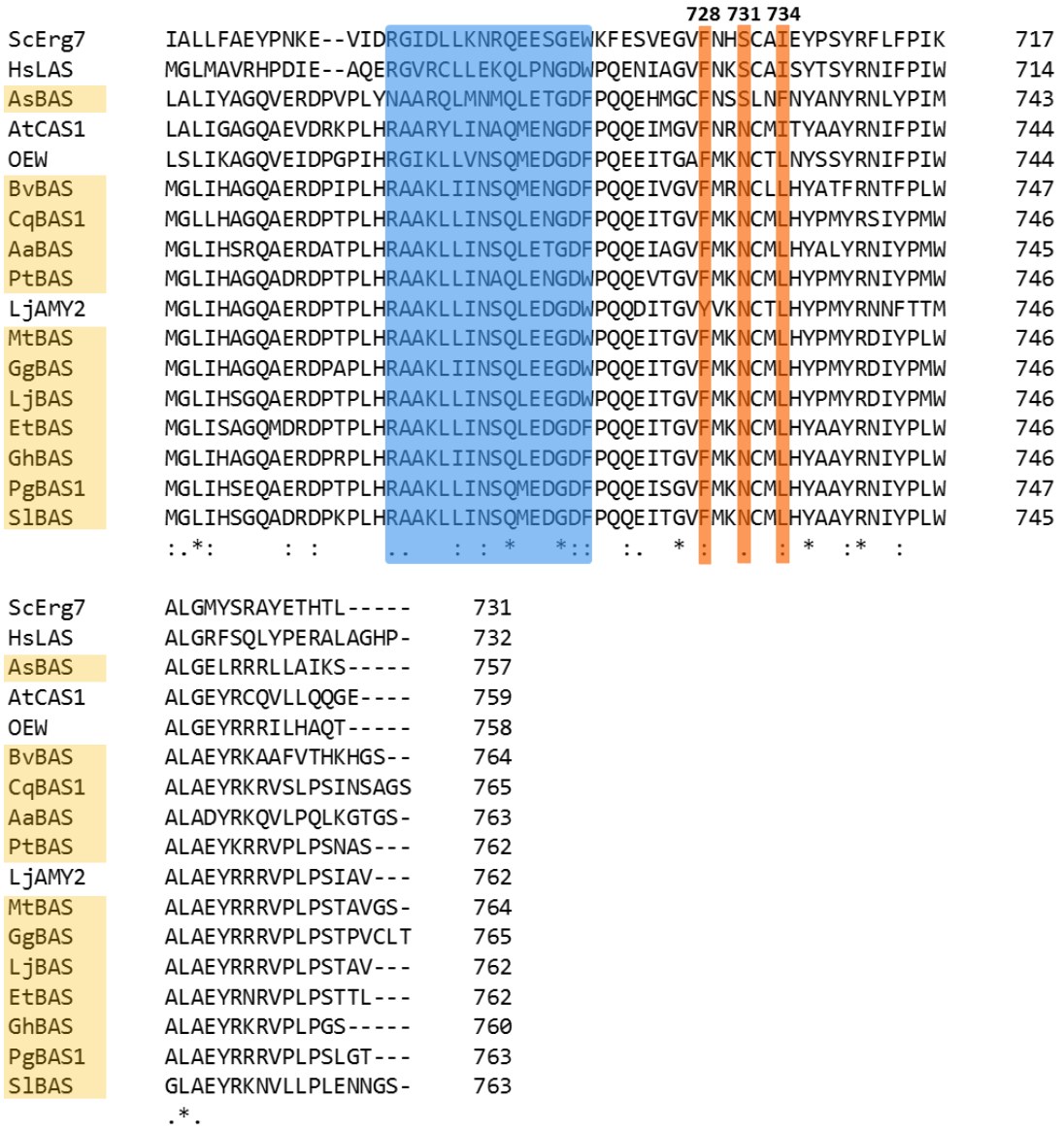


Figure 4.5. Multiple sequence alignment of β-amyrin synthases and other OSCs. Protein sequences were aligned using Clustal Omega (www.ebi.ac.uk/Tools/msa/clustalo/) (Sievers and Higgins, 2018) with default parameters. Conserved sequences are highlighted. Red: DCTAE motif and hydrogen-bonded Cys; green: MWCYCR motif; blue: QW motifs; orange: other conserved residues (numbered relative to EtBAS). The concensus is shown under the alignment: identical residues are marked by '*', conservative substitutions by '.', and semi-conservative substitutions by ':'. The BASs tested in the present study are highlighted (see Table 4.1 for accession numbers). ScErg7: *S. cerevisiae* lanosterol synthase (NP_011939); HsLAS: *Homo sapiens* lanosterol synthase (AAC50184); AtCAS1: *Arabidopsis thaliana* cycloartenol synthase (NP_178722); OEW: *Olea europaea* lupeol synthase (BAA86930); LjAMY2: *Lotus japonicus* multifunctional lupeol/β-amyrin synthase. The sequences of AaBAS, BvBAS, MtBAS and SlBAS contained an additional 'GS' sequence at the C-terminus due to the modular DNA assembly platform used (MoClo-YTK) (Lee et al., 2015) but were otherwise identical to those from GenBank. The CqBAS1 sequence included this C-terminal 'GS' and had a D196Y mutation (highlighted dark red; see main text).

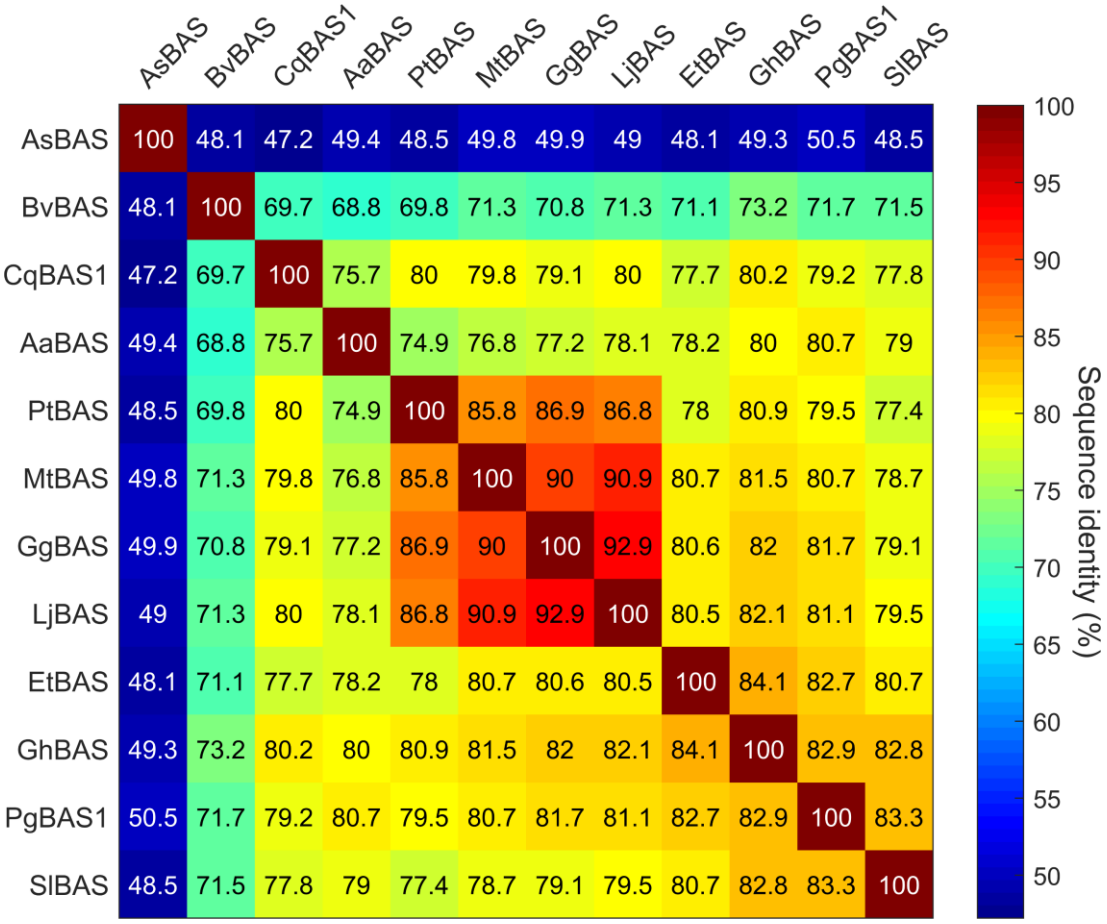


Figure 4.6. Percent identity matrix of BAS sequence identities. Protein sequences were aligned using ClustalOmega (www.ebi.ac.uk/Tools/msa/clustalo/) (Sievers and Higgins, 2018) with default parameters to obtain a percent identity matrix (PIM). All proteins except BvBAS and AsBAS had sequence identities > 75 %. For accession numbers see Table 4.1.

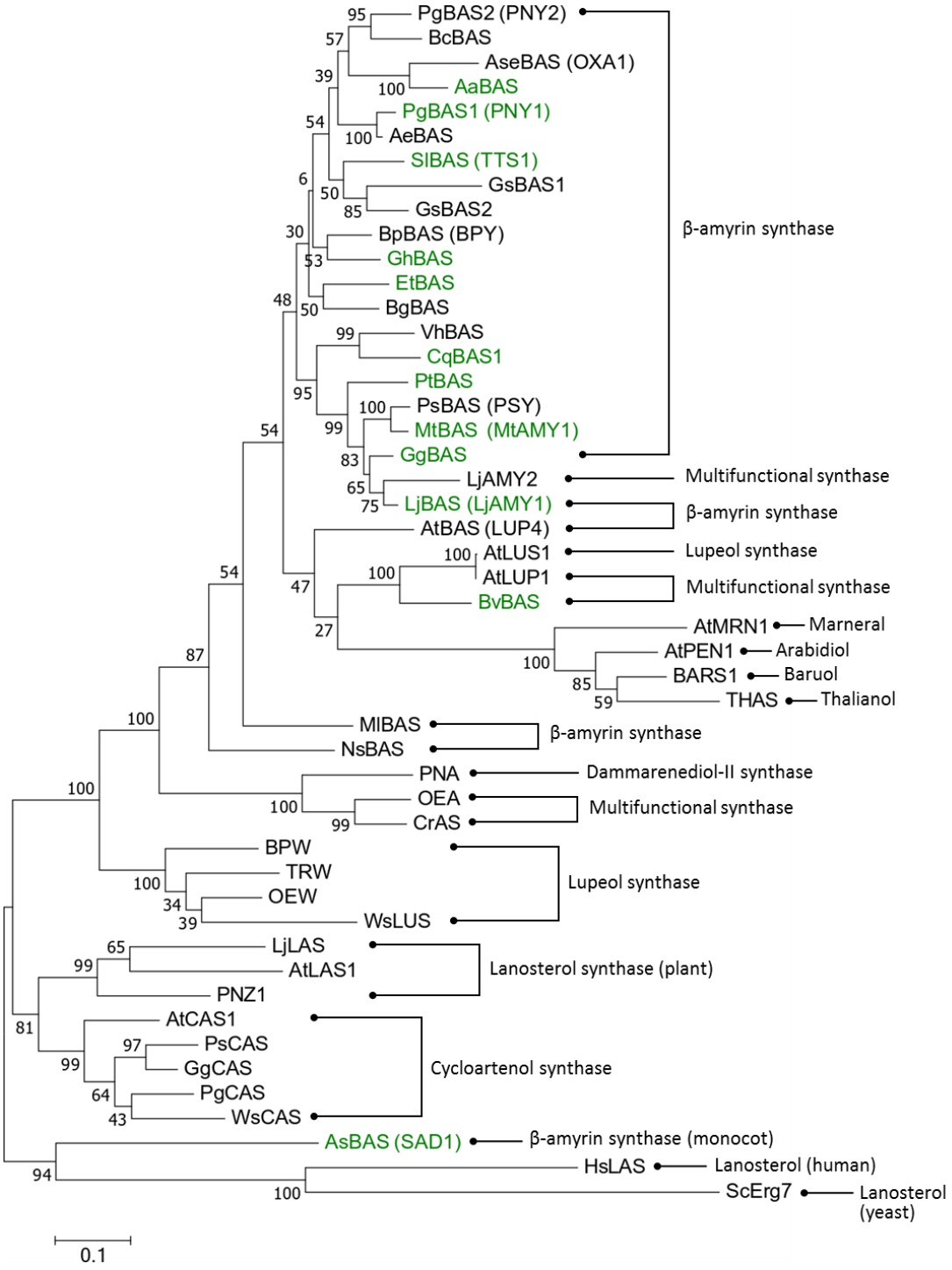

PgBAS2 (PNY2)
BcBAS
AseBAS (OXA1)
AaBAS
PgBAS1 (PNY1)
AeBAS
SlBAS (TTS1)
GsBAS1
GsBAS2
BpBAS (BPY)
GhBAS
EtBAS
BgBAS
VhBAS
CqBAS1
PtBAS
PsBAS (PSY)
MtBAS (MtAMY1)
GgBAS
β-amyrin synthase
LjAMY2
Multifunctional synthase
LjBAS (LjAMY1)
β-amyrin synthase
AtBAS (LUP4)
AtLUS1
Lupeol synthase
AtLUP1
Multifunctional synthase
BvBAS
AtMRN1
Marneral
AtPEN1
Arabidiol
BARS1
Baruol
THAS
Thalianol
MlBAS
NsBAS
β-amyrin synthase
PNA
Dammarenediol-II synthase
OEA
CrAS
Multifunctional synthase
BPW
TRW
OEW
WsLUS
Lupeol synthase
LjLAS
AtLAS1
PNZ1
Lanosterol synthase (plant)
AtCAS1
PsCAS
GgCAS
PgCAS
WsCAS
Cycloartenol synthase
AsBAS (SAD1)
β-amyrin synthase (monocot)
HsLAS
Lanosterol (human)
ScErg7
Lanosterol (yeast)
0.1

Figure 4.7. Phylogenetic analysis of β-amyrin synthases (BASs) and other oxidosqualene cyclases (OSCs). The phylogenetic tree was generated from protein sequences by using the maximum likelihood method and bootstrapping for 1000 replicates. Bootstrap values (%) are displayed at each node. The BASs tested in the present study are highlighted green (see Table 4.1 for accession numbers), and the function of each OSC is annotated. The scale bar represents the number of substitutions per site. AtBAS (LUP4): *Arabidopsis thaliana* β-amyrin synthase (NP_178016.2); AeBAS: *Aralia elata* β-amyrin synthase (ADK12003.1); AseBAS (OXA1): *Aster sedifolius* β-amyrin synthase (AAX14716.1); BcBAS: *Bupleurum chinense* β-amyrin synthase (ABY90140.2); BgBAS: *Bruguiera gymnorrhiza* β-amyrin synthase (BAF80443.1); BpBAS (BPY): *Betula platyphylla* β-amyrin synthase (BAB83088.1); GsBAS1: *Gentiana straminea* β-amyrin synthase (ACO24697.1); GsBAS2: *G. straminea* β-amyrin synthase (AHX97777.1); MlBAS: *Maesa lanceolata* β-amyrin synthase (AHF49822.1); NsBAS: *Nigella sativa* β-amyrin synthase (ACH88048.1); PgBAS2 (PNY2): *Panax ginseng* β-amyrin synthase (BAA33722.1); PsBAS (PSY): *Pisum sativum* β-amyrin synthase (BAA97558.1); VhBAS: *Vaccaria hispanica* β-amyrin synthase (ABK76265.1); AtLUP1: *A. thaliana* Multifunctional triterpene synthase (NP_178018.1); LjAMY2: *Lotus japonicus* Multifunctional β-amyrin/lupeol synthase (AAO33580.1); CrAS: *Catharanthus roseus* Multifunctional a/β-amyrin synthase (AFJ19235.1); OEA: *Olea europaea* Multifunctional triterpene synthase (BAF63702.1); AtLUS1: *A. thaliana* Lupeol synthase (AAD05032.1); BpLUS (BPW): *Betula platyphylla* Lupeol synthase (BAB83087.1); ToLUS (TRW): *Taraxacum officinale* Lupeol synthase (BAA86932.1); OeLUS (OEW): *Olea europaea* Lupeol synthase (BAA86930.1); WsLUS: *Withania somnifera* Lupeol synthase (AGA17939.1); PNA: *P. ginseng* Dammarenediol-II synthase (BAF33291.1); BARS1: *A. thaliana* Baruol synthase (NP_193272.1); AtPEN1: *A. thaliana* Arabidiol synthase (NP_567462.1); AtMRN1: *A. thaliana* Marneral synthase (NP_199074.1); THAS: *A. thaliana* Thalianol synthase (NP_001078733.1); AtCAS1: *A. thaliana* Cycloartenol synthase (NP_178722.1); PsCAS: *Pisum sativum* Cycloartenol synthase (BAA23533.1); PgCAS: *P. ginseng* Cycloartenol synthase (BAA33460.1); GgCAS: *Glycyrrhiza glabra* Cycloartenol synthase (BAA76902.1); WsCAS: *Withania somnifera* Cycloartenol synthase (ADG60271.1); AtLAS1: *A. thaliana* Lanosterol synthase (NP_001327790.1); LjLAS: *L. japonicus* Lanosterol synthase (BAE95410.1); PNZ1: *P. ginseng* Lanosterol synthase (BAA33462.1); HsLAS: *Homo sapiens* Lanosterol synthase (AAC50184.1); ScErg7: *Saccharomyces cerevisiae* Lanosterol synthase (NP_011939.2).

The BASs were codon-optimised for *S. cerevisiae*, synthesised, assembled into a low copy plasmid (CEN6/ARS4) under the control of the *GAL1* promoter, and transformed into *S. cerevisiae* BY4741 (MD-BAX series, see Table 4.1). For clarity, the strains will be referred to by the name of the BAS, and not by their strain code. All BASs were codon-optimised for *S. cerevisiae* except *AsBAS*, which failed gene synthesis and a wild-type sequence was used instead (provided by Thomas Louveau, Anne Osbourn Lab, John Innes Centre). Initially, correct clones of *CqBAS1* could not be obtained, and sequencing revealed that the *E. coli* Tn5 transposon had been inserted into the open reading frame. The transposon was removed by PCR and silent mutations were introduced to prevent its reinsertion, but all resulting clones contained unintended missense mutations. A clone with a D196Y mutation was selected for analysis, as this residue is not predicted to be involved in catalysis and has little sequence conservation (Figure 4.5).

The resulting strains and MD-N1 were analysed for β-amyrin production by GC-MS as described before. Ergosterol titres and the OD_{600} at the end of the production run (the "final OD_{600}") were also measured, as they may be affected by BAS expression. In addition, growth in both glucose- and galactose-supplemented media was monitored using a 96-well microplate assay to determine the effects of BAS expression on growth rate. The glucose precultures used for the production runs were used to inoculate media in the 96-well plate, and OD_{595} was read every 15 min for 96 h in a Tecan Sunrise plate reader. The experiment was repeated three times, each with three biological replicates, for a total of nine replicates.

β-amyrin production varied (Figure 4.8 A). The AaBAS and CqBAS1 strains had the highest β-amyrin titres at 10.6 mg/L, while SlBAS and GhBAS produced less at 3.7 mg/L and 5.9 mg/L, respectively. Notably, PgBAS1 produced very low quantities of β-amyrin that fell below the linear region of the standard curve, preventing reliable calculation of titres. The standard curve can be used to calculate β-amyrin titres down to 1.7 mg/L (see Chapter 3), indicating that PgBAS1 produced < 1.7 mg/L β-amyrin. Furthermore, the normalised β-amyrin peak area for PgBAS1 was 97-fold lower than for AaBAS and CqBAS1 (see Figure 4.14 for the TICs). The final OD_{600} also varied, and negatively correlated with the β-amyrin titre (Figure 4.8 C, Figure 4.9 A). The high β-amyrin producers AaBAS and CqBAS1 reached an OD_{600} of 4.7 and 4.4, respectively, while the low producers PgBAS1, GhBAS and SlBAS reached an

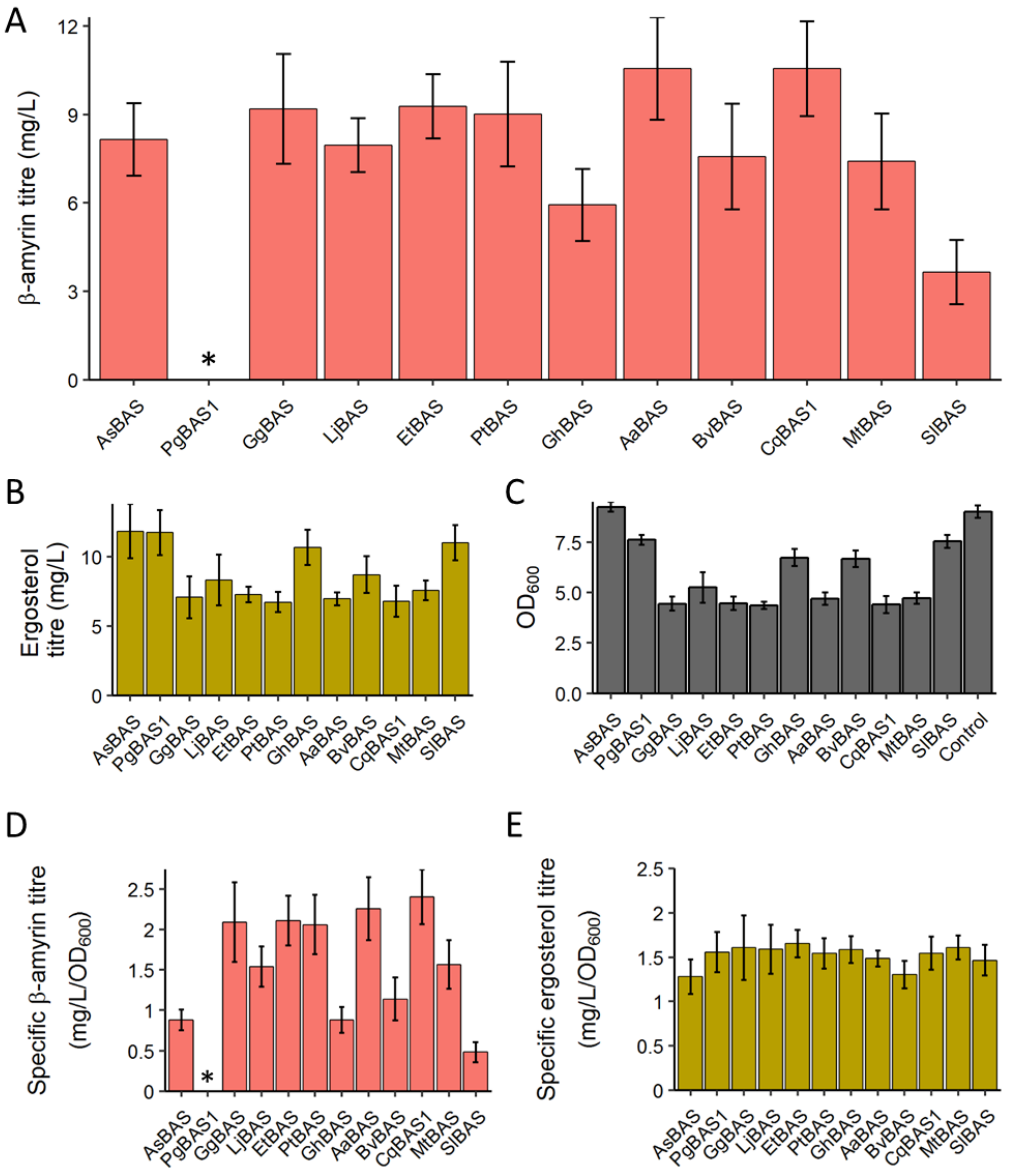


Figure 4.8. Comparison of β-amyrin synthase productivity in yeast. (A-B) β-amyrin and ergosterol titres as measured by GC-MS. **(C)** OD_{600} measured at the end of the production run (96 h). Titres were divided by the final OD_{600} to obtain specific titres (mg/L/OD_{600}) for β-amyrin **(D)** and ergosterol **(E)**. Strains produced varying amounts of β-amyrin and ergosterol and reached different cell densities dependent on the BAS. Specific ergosterol titres were fairly consistent across all strains, while specific β-amyrin titres showed large variations. Control: MD-N1 (empty vector). *, titre could not be calculated as the peak area fell below the linear region of the standard curve. Shown is the mean ± 1 standard deviation for 3 biological replicates repeated across three experiments (n = 9).

OD_{600} of 6.7 - 7.6. By comparison, MD-N1 reached an OD_{600} of 9.0. A striking exception to this correlation was AsBAS, which reached an OD_{600} of 9.2 yet produced a relatively large amount of β-amyrin (8.2 mg/L) (Figure 4.9 A).

Specific β-amyrin titres were calculated by dividing the titre by the final OD_{600} (mg/L/OD_{600}). These showed a similar overall pattern to the total β-amyrin titres (Figure 4.8 D), although specific titres for AsBAS, GhBAS, BvBAS and SlBAS were particularly low because of their high OD_{600}. Thus, while total β-amyrin production was relatively high for AsBAS (8.2 mg/L), specific production was low (0.88 mg/L/OD_{600}). Similar to total β-amyrin titres, there was also a negative correlation between specific β-amyrin titres and final OD_{600} (Figure 4.9 B) with AsBAS having a higher specific β-amyrin titre than would be expected for its OD_{600} (Figure 4.9 B). Overall, it appears that BAS expression slows growth, and that the magnitude of this effect increases with β-amyrin production.

Ergosterol titres also varied (Figure 4.8 B), showing a positive correlation with the final OD_{600} (Figure 4.9 C) and a negative correlation with β-amyrin titre (Figure 4.9 E). PgBAS1, GhBAS and SlBAS, which produced relatively low amounts of β-amyrin (≤ 5.9 mg/L), had the highest ergosterol titres of the BAS strains, ranging from 10.7 mg/L to 11.8 mg/L. AaBAS and CqBAS1, which had the highest β-amyrin titres, accumulated less ergosterol at 7.0 mg/L and 6.8 mg/L, respectively. The ergosterol titre could not be accurately calculated for MD-N1 because three replicates lied above the linear region of the standard curve. When excluding these replicates the titre was calculated as 14.7 mg/L, however this is an underestimate due to the exclusion of the replicates with greater peak areas. In addition to reaching a similar OD_{600} to MD-N1, AsBAS had a high ergosterol titre (11.9 mg/L) despite producing a relatively high amount of β-amyrin (8.2 mg/L) (Figure 4.9 E). In general, strains expressing BASs reached a lower OD_{600} and accumulated less total ergosterol than MD-N1, with the poorer β-amyrin producers (PgBAS1, GhBAS and SlBAS) reaching higher OD_{600} and accumulating more total ergosterol.

Specific ergosterol titres were similar for all strains (Figure 4.8 E) and did not vary with final OD_{600} or specific β-amyrin titre (Figure 4.9 D, F). Because both ergosterol and β-amyrin are derived from oxidosqualene, it might be expected that β-amyrin production would redirect metabolic flux from ergosterol to β-amyrin biosynthesis, reducing

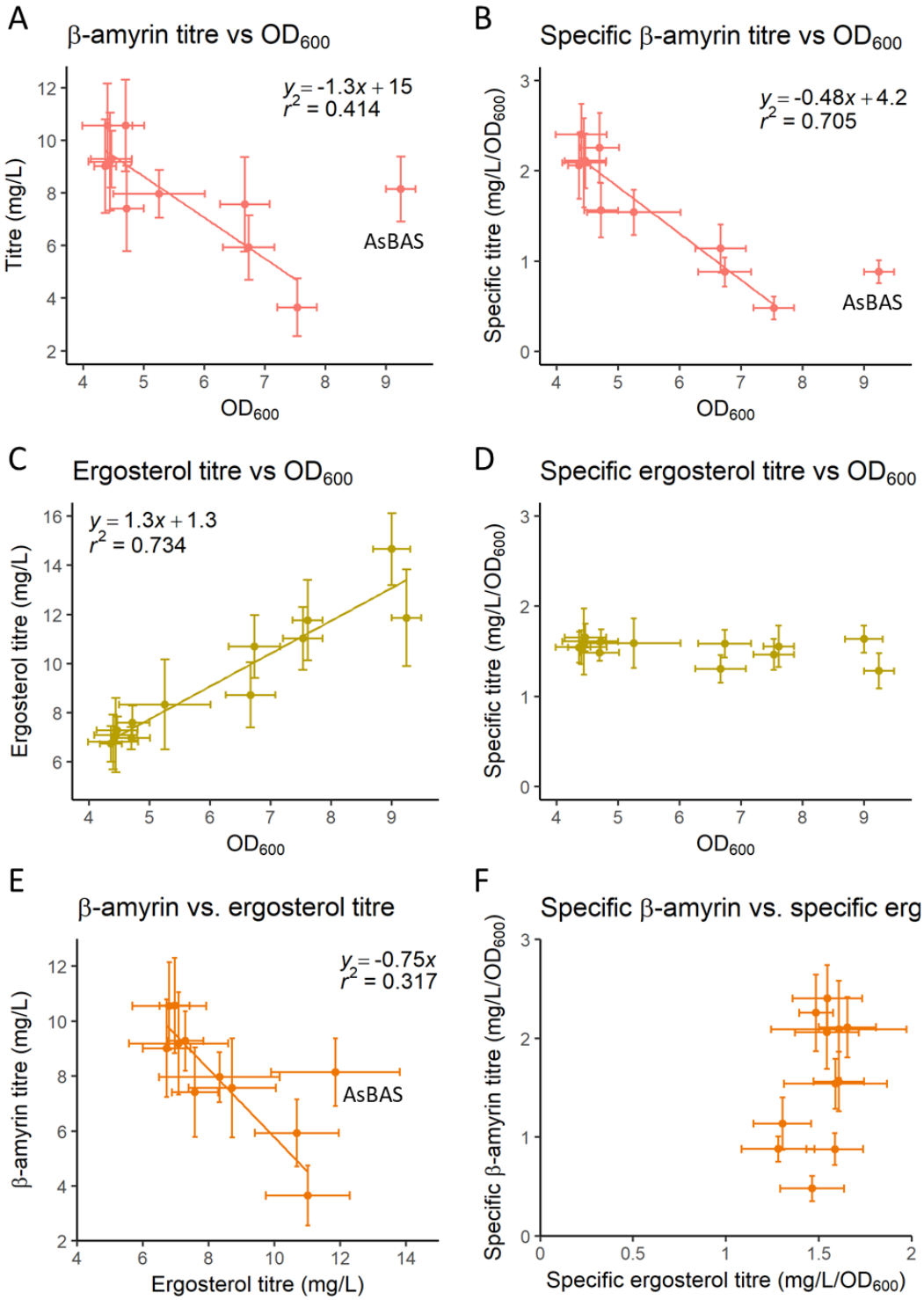


Figure 4.9. Correlations between total and specific titres and the final OD_{600}. Correlations between total **(A)** or specific **(B)** β-amyrin titres and OD_{600}, total **(C)** or specific **(D)** ergosterol titres and OD_{600}, β-amyrin and ergosterol titres **(E)**, and specific β-amyrin and ergosterol titres **(F)**. Shown is the mean ± 1 standard deviation for 3 biological replicates repeated across three experiments (n=9). Lines were fit using R and AsBAS was treated as an outlier for plots A, B and E.

cellular ergosterol levels. However, the lack of variation in specific ergosterol titres indicates that cellular ergosterol levels were similar for all strains, and that BAS expression had little or no effect on ergosterol biosynthesis. If flux was being redirected to β-amyrin, we would expect an inverse correlation between specific ergosterol titres and specific β-amyrin titres, as increased β-amyrin production would reduce ergosterol levels. The fact that this was not observed could be because β-amyrin production was relatively low and that Erg7p, the yeast enzyme that cyclises oxidosqualene to lanosterol as part of ergosterol biosynthesis, is known to be highly active (Veen et al., 2003). In addition, the ergosterol biosynthetic pathway is tightly regulated, which may allow cells to maintain stable amounts of ergosterol under a range of conditions (Klug and Daum, 2014; Hu et al., 2017).

The lack of variation in specific ergosterol titres, combined with the positive correlation between total ergosterol and OD_{600}, indicate that the observed variations in ergosterol titre are due to differences in growth, rather than depletion of cellular ergosterol levels caused by BAS expression. It therefore appears that BAS expression reduces growth rate and this in turn reduces total ergosterol titres, explaining the observed correlations between β-amyrin titres, ergosterol titres and OD_{600}.

In general, the growth curves in galactose-supplemented media corroborated the OD_{600} measurements from the production runs (Figure 4.10 A). It should be noted that the plate reader gives different OD values to the spectrophotometer that was used to measure the OD_{600} of the production run cultures. As such, relative strain growth can be compared but the absolute OD values cannot. When grown in galactose, most BAS strains showed a considerable reduction in growth rate and an increased lag phase compared with MD-N1, while AsBAS grew at almost the same rate, indicating that its expression had little effect on cell growth (Figure 4.10 A). This is particularly striking compared to the slow growth rates of most of the other BAS strains, and is in agreement with the OD_{600} data from the production runs. In addition to AsBAS, the strains expressing PgBAS1 and BvBAS grew noticeably faster than the others, although not as fast as AsBAS. While the growth curves generally agreed with the OD_{600} measurements from the production runs, this was not the case for SlBAS. This strain grew very slowly in the plate reader assay and reached the lowest OD_{595} of the strains tested, while in the production run it reached a relatively high OD_{600}. However,

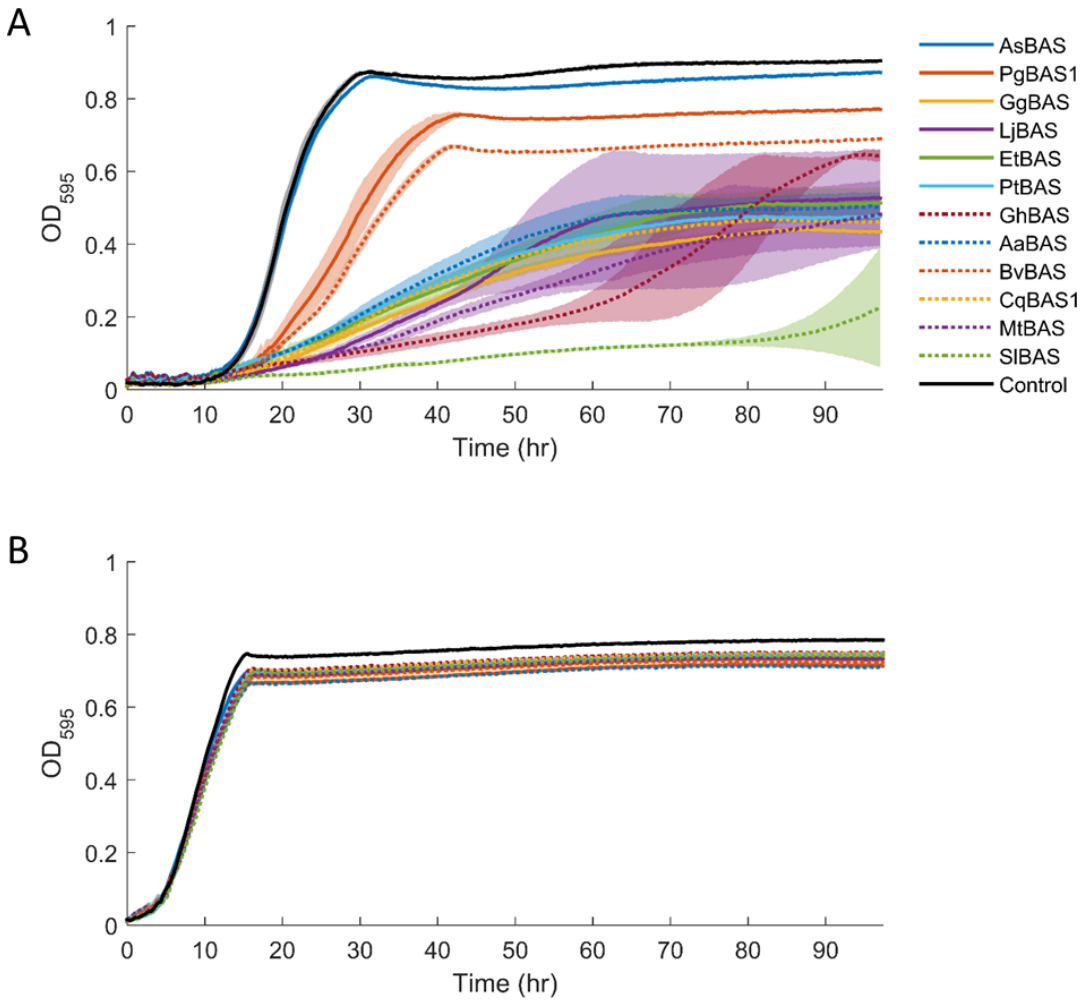


Figure 4.10. Growth of BAS strains in a 96-well plate reader assay. Cells were grown in media supplemented with either galactose **(A)** or glucose **(B)**. OD_{595} was monitored using a Tecan Sunrise plate reader, reading every 15 min for 96 h. A variety of growth rates were observed in the galactose-supplemented media, while all strains grew similarly in glucose, with the empty vector control strain (MD-N1) growing slightly faster. Shown is the mean ± 1 standard deviation for 3 biological replicates from one experiment.

for the remaining strains, the relative OD reached by the end of the growth curve agreed well with those from the production run.

In contrast with the galactose growth curves, all strains grew similarly in glucose-supplemented media (Figure 4.10 B). MD-N1 had a long lag phase of approximately 10 h when grown in galactose, whereas this lag phase was much shorter (~ 4 h) when grown in glucose. The cells are first grown in glucose-supplemented media before being diluted into fresh glucose- or galactose-supplemented media for the plate reader assay. This sudden change in carbon source could explain the longer lag phase in galactose media, together with the fact that galactose is a poorer carbon source for yeast than glucose and that galactose metabolism is repressed in the presence of glucose and is activated only upon glucose depletion and in the presence of galactose (Timson, 2007). The fact that all strains grew similarly in glucose media might be explained by there being little BAS expression under these conditions. Indeed, western blots of selected BASs could detect protein after growth in galactose but not glucose (Figure 4.11 A). The exception was AsBAS, which was not detected after growth in galactose, although RT-PCR confirmed that the AsBAS gene was being expressed (Figure 4.11 B). The inability to detect AsBAS by western blot could be because the His-tag used for detection was inaccessible, or because it was expressed at low levels and was below the limit of detection.

The reduced growth rates of most BAS strains is quite striking. Some possible reasons are expression of the BAS, potential toxicity of β-amyrin, depletion of native metabolites such as ergosterol, the formation of toxic side-products by the BAS, or misfolding of the BAS leading to a burden on the unfolded protein response in yeast, thus affecting growth. As reasoned above, BAS expression does not appear to have depleted cellular ergosterol levels. However, depletion of other metabolites and cofactors were not monitored and so cannot be ruled out. Differences in BAS expression could account for the variation in β-amyrin production as well as the differences in growth. Lower expression would be expected to favour growth and decrease β-amyrin titres, and would be consistent with the negative correlation between β-amyrin titres and OD_{600}. Furthermore, the *GAL1* promoter from which the BASs are expressed gives very high expression levels, comparable to the constitutive *TDH3* promoter, which is among the strongest yeast promoters (Lee et al., 2015), and expression from this promoter would be expected to cause burden. It is possible that

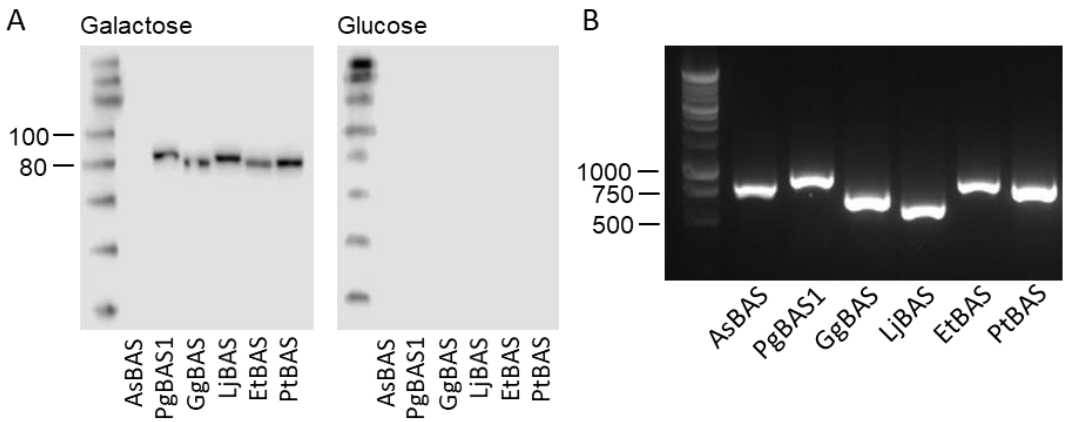


Figure 4.11. Analysis of BAS expression by western blot and RT-PCR. Western blots **(A)** and RT-PCR **(B)** were performed on strains expressing 6xHis-tagged AsBAS, PgBAS1, GgBAS, LjBAS, EtBAS and PtBAS. For the western blots, strains were grown in galactose (left) or glucose (right), while RT-PCR was performed on strains grown in galactose. All BASs but AsBAS were detected by the western blot, while mRNA was detected for all six BASs by RT-PCR. The expected molecular weights of the BASs were ~ 87 kDa. Molecular weights for the protein and DNA ladders are shown in kDa and bp, respectively.

AsBAS is expressed at very low levels, resulting in very similar growth to MD-N1 and the inability to detect the protein by western blot. Although it was also expressed from the strong *GAL1* promoter, poor translation efficiency could nevertheless cause it to be expressed poorly. If this were the case, AsBAS would be a highly active enzyme, as it produced a relatively large amount of β-amyrin. However, expression levels of the BASs would need to be determined to test this hypothesis.

4.4 BvBAS is a multifunctional OSC

The total ion chromatogram of BvBAS revealed two additional peaks to β-amyrin, eluting at ~ 17.7 min (peak U1) and 19.1 min (U2), which from their mass spectra appeared to be triterpenes (Figure 4.12). Although the BASs tested here were supposedly monofunctional (producing only β-amyrin), there are numerous multifunctional OSCs that produce multiple triterpenes, often including β-amyrin, α-amyrin, and/or lupeol (Ghosh, 2016). Alpha-amyrin has a very similar structure to β-amyrin, differing only in the position of a methyl group, while lupeol is a pentacyclic triterpene like β-amyrin but with a five-membered E-ring instead of a six-membered E-ring (Figure 4.13) (Thimmappa et al., 2014). These three triterpenes are derived from sequentially-formed carbocation intermediates during the cyclisation reaction

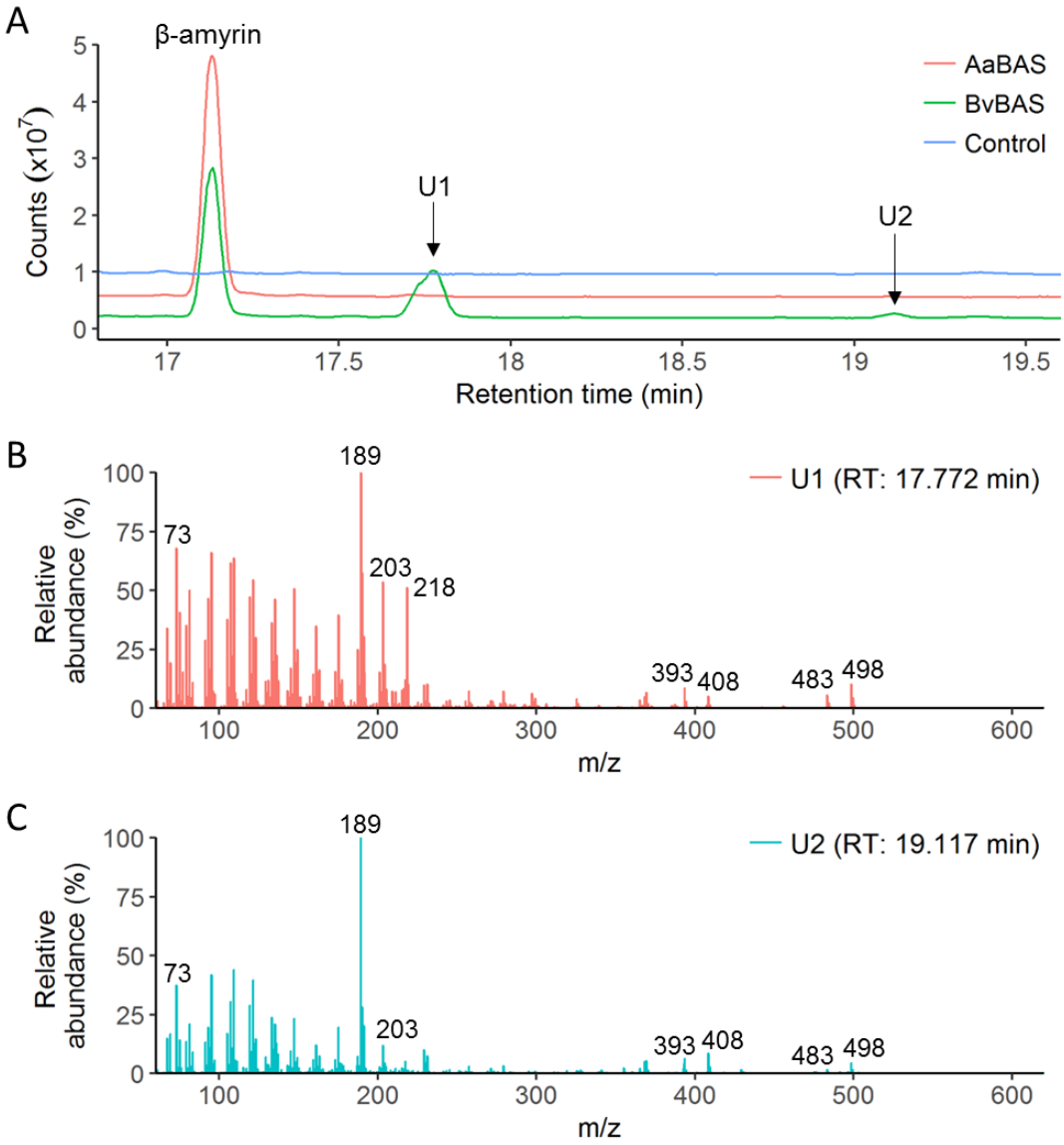


Figure 4.12. BvBAS is a multifunctional OSC producing multiple triterpenes. (A) Total ion chromatogram (TIC) overlay of the strains expressing AaBAS, BvBAS and of the MD-N1 control strain, showing two additional peaks (U1 and U2) observed for BvBAS. Mass spectra of peaks U1 **(B)** and U2 **(C)**, with background spectra (corresponding to 18.104 - 18.398 min of each sample) subtracted. Peak U1 may correspond to lupeol based on comparison with a published mass spectrum of this triterpene (see Fig. S3 from Khakimov *et al.*, 2015). The identity of U2 is unknown but its mass spectrum is consistent with it being a triterpene.

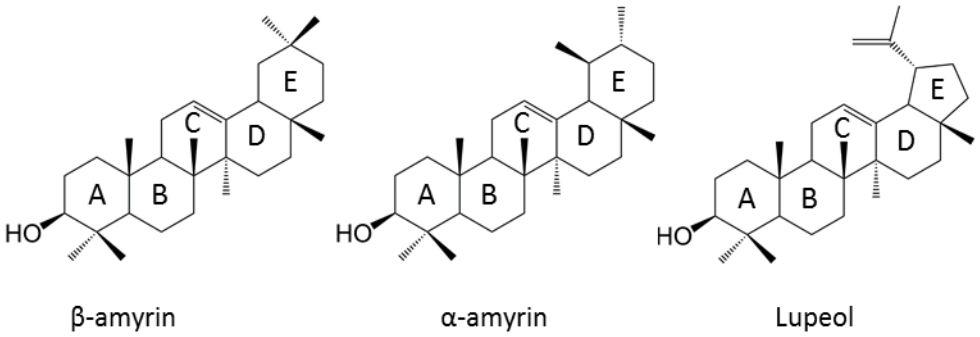


Figure 4.13. **Chemical structures of β-amyrin, α-amyrin and lupeol.** The rings are labelled A through E. β-amyrin and α-amyrin have a 6 membered E-ring while lupeol has a five membered E-ring, and β-amyrin and α-amyrin differ in the position of a methyl group at the E-ring.

(the lupanyl, germanicyl and oleanyl cations for lupeol, α-amyrin and β-amyrin, respectively; see Figure 1.3) (Thimmappa et al., 2014). It is therefore possible that these additional peaks could correspond to α-amyrin and/or lupeol.

The mass spectrum of U1 is consistent with lupeol mass spectra reported in the literature (e.g. see Fig. S3 from Khakimov *et al.*, 2015). In addition, the sequence of BvBAS is consistent with it producing a mixture of triterpenes. Mutational studies of PgBAS1 have shown that a W259L mutation in the MWCYCR motif, which is highly conserved among BASs, converts it into a multifunctional OSC that produces a mixture of lupeol, β-amyrin, and small quantities of other triterpenes (Kushiro et al., 2000). BvBAS contains this mutation, having TLCYCR instead of MWCYCR (Figure 4.5). Lupeol synthases (LUSs) have MLCYCR instead of MWCYCR, and a L256W mutation in the MLCYCR motif of a LUS from *Olea europaea* caused it to produce mainly β-amyrin, with only small amounts of lupeol observed (Kushiro et al., 2000). Thus, the second residue of the MWCYCR motif is critical for product specificity, and the presence of a leucine at this position in BvBAS indicates that it is likely to produce lupeol. Although BvBAS also differs in the first residue of the MWCYCR motif, having Thr instead of Met, this may not be important for product specificity. A PgBAS1 mutant containing an MW258IL mutation (having ILCYCR) had a very similar product profile to the W259L mutant (having MLCYCR) (Kushiro et al., 2000).

The identity of peak U2 is unknown, but its mass spectrum is consistent with it being a triterpene. It contains ions with m/z values of 498, 483, 408 and 393, which would

be consistent with a triterpene parent ion with a single TMS group (m/z 498) and fragment ions resulting from the loss of a methyl and/or a TMSOH group from this parent ion. Furthermore, the presence of peaks with m/z values of 189 and 203 are consistent with rDA fragmentation of a triterpene. Indeed, these peaks are seen in the mass spectra of both β-amyrin (e.g. see Figure 4.4 B) and lupeol (see Fig. S3 of Khakimov *et al.*, 2015 and the putative lupeol mass spectrum in Figure 4.12 B). Kushiro *et al.* also reported small amounts of butyrospermol and germanicol for the PgBAS1 W269L mutant, in addition to β-amyrin and lupeol (Kushiro et al., 2000). These are therefore possible candidate identities for U2. Thus, BvBAS is a multifunctional triterpene synthase, mainly producing β-amyrin, low amounts of most likely lupeol, and an unknown triterpene.

Because BvBAS produced additional triterpenes, the other BASs were reanalysed to determine if they also produced additional products. AsBAS and EtBAS had clear yet small peaks with a similar retention time to peak U1 of BvBAS, at ~ 17.7 min (Figure 4.14). The other BASs also had very small putative peaks with similar retention times, but the signal was too weak to determine whether they were genuine peaks or due to noise. PgBAS1 had no peak at this position, which might be explained by its low activity, preventing any trace products from being detected. Due to the low intensity of the peaks, background spectra (~18.6-19.0 min of each sample) were subtracted to facilitate analysis of the mass spectra. In general, the mass spectra appeared consistent with lupeol, but the low abundance made some difficult to interpret (Figure 4.14). To confirm lupeol production, a splitless injection would be required to generate enough signal to reliably interpret the mass spectra. Lupeol has been reported as a co-product of β-amyrin for several OSCs (Ghosh, 2016), and as discussed above, a single mutation is enough to convert a BAS into a mixed β-amyrin/lupeol synthase (Kushiro et al., 2000).

It is not surprising that the other BASs may produce additional triterpenes, despite most being previously characterised as monofunctional. The cyclisation reaction that converts oxidosqualene to a triterpene occurs through many intermediates, and termination of the reaction at any one intermediate results in the generation of a corresponding triterpene (see section 1.1.2 and Figure 1.3). There are therefore many possible triterpenes that can be produced through this reaction. Indeed, approximately 150 OSC cyclisation products have been reported to date (Hoshino,

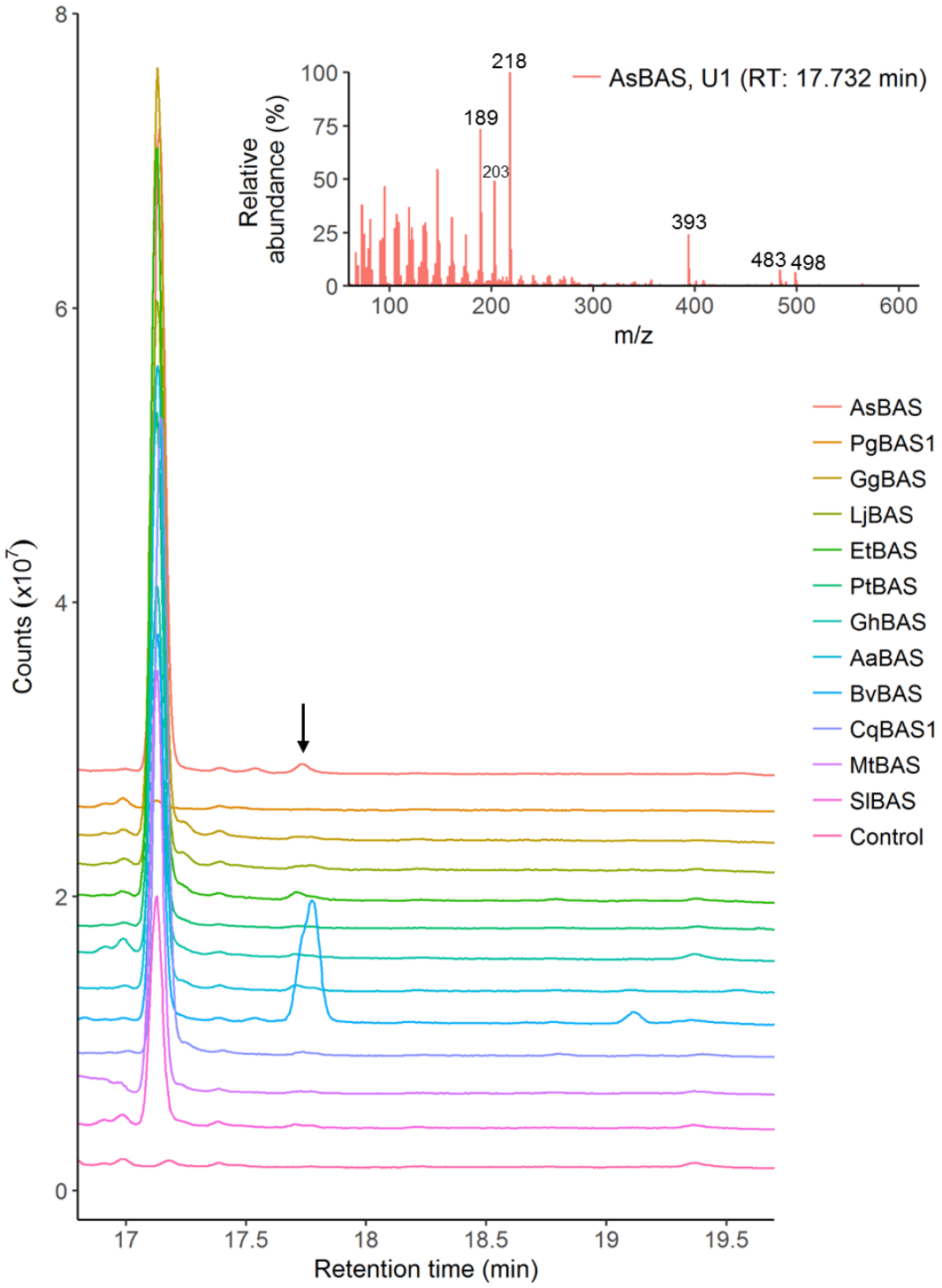


Figure 4.14. Additional triterpenes produced by the BASs. Analysis of TICs revealed that some BASs also had peaks corresponding to U1 from BvBAS in low abundance. Peaks corresponding to U2 were not observed. Inset: mass spectrum of peak U1 from AsBAS, with a background spectrum (corresponding to 18.104 – 18.398 min of the sample) subtracted. Peak U1 may correspond to lupeol.

2017). After cyclisation is initiated by the OSC (through protonation of the epoxide of oxidosqualene), the reaction occurs spontaneously, with the OSC acting to block unwanted cyclisation pathways to favour the production of the desired triterpene(s) (Thimmappa et al., 2014). Thus, the cyclisation reaction intrinsically acts to generate a diverse range of triterpenes, and the OSC ensures that one or a few products are favoured (Thimmappa et al., 2014). We would therefore expect monofunctional OSCs, such as the BASs tested here, to produce small amounts of other triterpenes in addition to the favoured one(s).

4.5 Effect of a catalytically inactive β-amyrin synthase on cell burden

The BAS strains displayed a variety of growth rates, with many growing slowly compared to MD-N1, indicating that BAS expression caused significant burden on the cells. AaBAS conferred high β-amyrin production and a slow growth rate, while AsBAS interestingly had little effect on growth rate yet still produced a relatively large amount of β-amyrin. To investigate the cause of this burden, a strain expressing a catalytically inactive variant of AaBAS was generated. It was previously reported that the D485N mutation in the DCTAE motif of EtBAS completely eliminated catalytic activity (Ito et al., 2013). This motif is highly conserved across all OSCs, and the Asp residue is responsible for initiating the cyclisation reaction by protonating the epoxide group of oxidosqualene (Hoshino, 2017). The equivalent mutation (D484N) was introduced into AaBAS to generate the catalytically inactive variant dAaBAS, which was transformed into yeast to generate strain MD-dBA14. This strain was cultured for GC-MS analysis and growth curves were generated as before. AaBAS, AsBAS and MD-N1 (the empty vector control) were also analysed for comparison.

No β-amyrin or other triterpenoids were observed for dAaBAS in the GC-MS profile, indicating that it was indeed catalytically inactive (Figure 4.15 C). Its TIC was very similar to MD-N1. By contrast, AaBAS had a distinct peak corresponding to β-amyrin. The growth curves showed that in galactose-supplemented media, dAaBAS grew at an intermediate rate between AaBAS and MD-N1, and that AsBAS grew similarly to MD-N1 as before (Figure 4.15 A). This indicates that the catalytic activity of AaBAS may indeed contribute to burden, but the fact that it grew slower than MD-N1 indicates that this is not the only cause. As with previous results, all strains grew similarly in glucose-supplemented media (Figure 4.15 B).

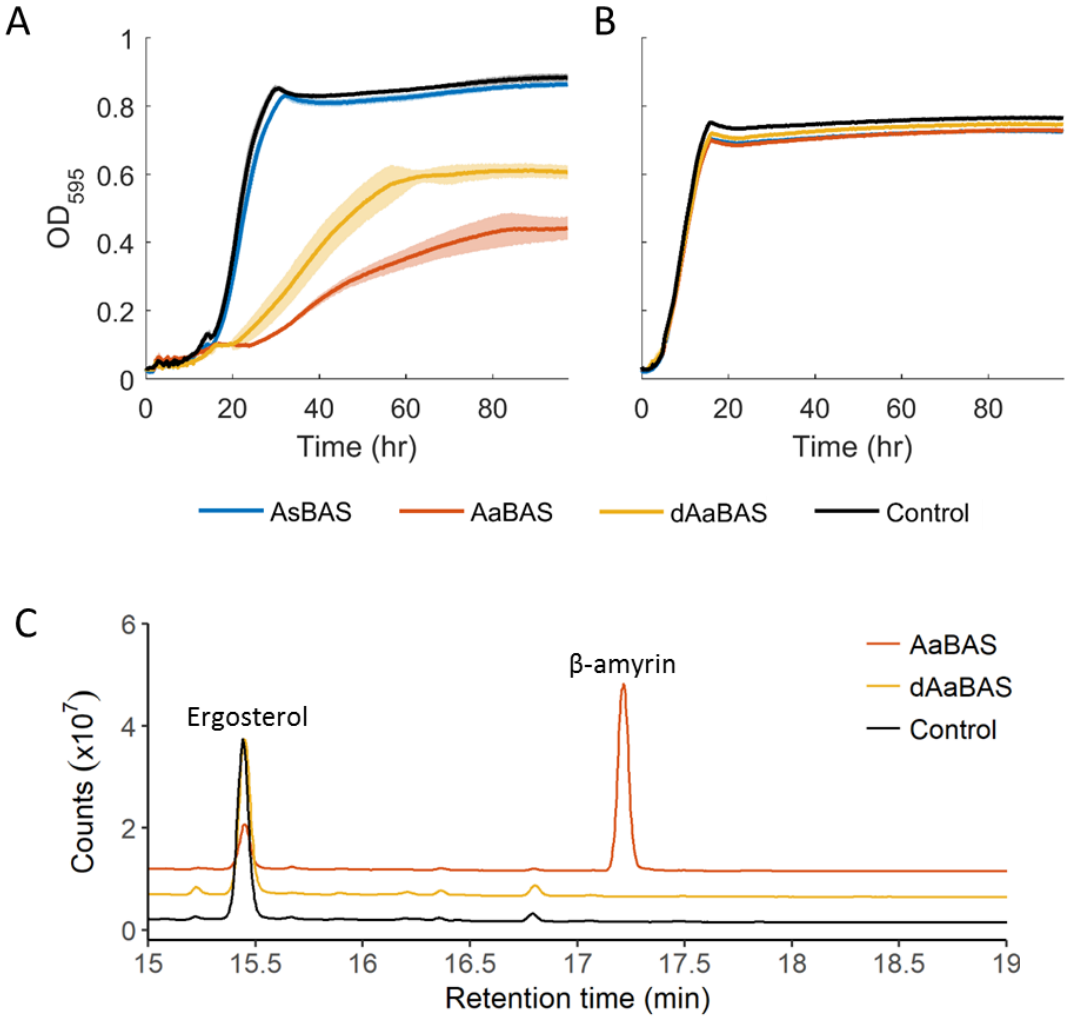


Figure 4.15. Effect on growth of a catalytically inactivated BAS. Strains expressing *AsBAS*, *AaBAS*, a catalytically inactive AaBAS (*dAaBAS*) and the control strain MD-N1 were grown in galactose **(A)** or glucose **(B)** supplemented media. Cells were grown in 96-well plates and OD_{595} was measured every 15 min for 96 h using a Tecan Sunrise plate reader. While all strains grew similarly in glucose, a variety of growth rates were observed in galactose, with dAaBAS growing at an intermediate rate between AsBAS/MD-N1 and AaBAS. Shown are the means of three biological replicates ± 1 standard deviation, except for dAaBAS, for which the mean of two biological replicates is shown. **(C)** Total ion chromatogram from GC-MS analysis showing that dAaBAS does not produce β-amyrin.

It is possible that the D484N mutation affected the expression level of *dAaBAS*. A reduced expression level would be expected to improve strain growth, and could explain the growth phenotype observed here. Indeed, Hoshino *et al.* report that the expression levels of *EtBAS* mutants were often lower than the wild-type enzyme (Hoshino, 2017). They recommended normalising the activity of each enzyme to its expression level to account for these differences, which can be done by simply dividing the amount of triterpene by the amount of enzyme (Hoshino, 2017). Unfortunately, it was not possible to accurately determine the expression levels of the BASs. Further work is therefore necessary to thoroughly investigate the burden caused by BAS expression.

4.6 *In silico* analysis of BASs to explain differences in activity

To further understand the differences in activity and burden caused by the BASs, *in silico* analyses were performed. The multiple sequence alignment (Figure 4.5) showed no obvious amino acid variants that might explain the differences in activity. The MWCYCR motif was highly conserved amongst all BASs except for BvBAS (see above) and AsBAS, which instead had FWCFTR. Kushiro *et al.* found that mutations in the MWCYCR affected the product profile of PgBAS1 (Kushiro et al., 2000). The W259L mutation caused PgBAS1 to produce significant amounts of lupeol, while the Y261H mutation caused it to produce predominantly dammaranes (tetracyclic triterpenes). Kushiro *et al.* did not measure the expression level of PgBAS1 and so their work does not provide information on the overall catalytic activity (in terms of total triterpenes produced) of the mutants (Kushiro et al., 2000). AsBAS was monofunctional, so of greater interest here are the amino acid variants that may affect its overall activity, rather than its product profile. Nevertheless, it is reasonable that the mutations identified by Kushiro *et al.* might affect the overall activity, so it is worth taking this data into account.

AsBAS has a Phe instead of a Tyr (**F**WC**FT**R instead of **M**WC**YC**R), however, these amino acids are very similar in structure, differing in only a hydroxyl group, and therefore may not have a significant effect on activity. Similarly, AsBAS has a Phe instead of a Met, both of which are hydrophobic. The MW258IL PgBAS1 mutant had a very similar product profile to the W259L mutant, indicating that a mutation at this position (at least to a similar amino acid) has little effect on the product ratio (Kushiro

et al., 2000). Lastly, AsBAS differs from the MWCYCR consensus at the fifth position, having a Thr instead of a Cys. However, a C262S mutation in PgBAS1 had no effect on the product ratio, and given the similarity of Ser to Thr, this suggests that the presence of a Thr in AsBAS may not affect activity.

As expected, the DCTAE motif was highly conserved amongst all BASs. As discussed above, the Asp initiates the cyclisation reaction by protonating the epoxide of oxidosqualene (Hoshino, 2017). This Asp is conserved across all OSCs and represents a common mechanism to initiate cyclisation. Two highly conserved Cys residues, one in the DCTAE motif and the other corresponding to C564 of EtBAS, are believed to hydrogen-bond with this Asp, increasing its acidity to allow it to protonate oxidosqualene (Thoma et al., 2004; Ito et al., 2013). Indeed, mutation of these residues to Ala in EtBAS (C486A and C564A mutants) reduced catalytic activity to 50 % and 2 % of the wild-type (Ito et al., 2013), indicating that C564 is the more important interaction partner.

F413 (EtBAS numbering) is highly conserved amongst the BASs and mutational studies on EtBAS suggest that its role is to provide steric bulk to prevent premature termination of the cyclisation cascade (Ito et al., 2017). AsBAS has a Tyr at this position, which due to its similar structure to Phe may fulfil a similar role. Similarly, F474 of EtBAS also provides steric bulk (Ito et al., 2014) and is present in all the BASs except AsBAS and BvBAS, which have a Leu instead. Like Phe, Leu is a hydrophobic amino acid but would not provide quite as much steric bulk as Phe. Nevertheless, AsBAS is monofunctional, indicating that premature cyclisation does not occur. F728 of EtBAS is believed to stabilise the baccarenyl and oleanyl carbocation intermediates of the cyclisation reaction (Hoshino, 2017), and is present in all of the BASs.

Notably, AsBAS has a Tyr at position 533, while the other BASs have a Trp at this position. Mutations of the equivalent residue in EtBAS (W534) considerably reduced β-amyrin production (< 10 % of wild-type activity). This included a W534Y mutation (Hoshino et al., 2017). Thus, it is possible that the presence of a Tyr instead of a Trp in AsBAS may lower its activity, and that a Y533W mutation in AsBAS could boost β-amyrin production. V483 of EtBAS is located close to W534, and mutations in V483 caused EtBAS to produce monocyclic triterpenes (Hoshino et al., 2017). However, V483 is conserved amongst all the BASs tested here. Mutants of EtBAS at W612 and L734 produced aberrant cyclisation products and had reduced activities (Aiba et al.,

2018). However, all BASs had this Trp and all but AsBAS had the Leu; AsBAS instead had a Phe. The purpose of L734 may be to provide steric bulk (Aiba et al., 2018) and Phe may be able to provide a similar role in AsBAS, although it is a slightly larger amino acid.

All of the BASs had a conserved Asn at position 731 (EtBAS numbering) except for AsBAS, which had a Ser (S728). It was previously shown that an S728F AsBAS mutant produced the tetracyclic triterpene dammarenediol-II when expressed in oat (*Avena strigosa*) (Salmon et al., 2016). When expressed in *S. cerevisiae* GIL77, this mutant produced mainly epoxydammara-3,25-diol, which is derived from the cyclisation of dioxidosqualene (Salmon et al., 2016). Oxidosqualene is produced by the epoxidation of squalene by squalene epoxidase (e.g. Erg9p in *S. cerevisiae*), and squalene epoxidase can oxidise oxidosqualene again to produce dioxidosqualene, which has two epoxide groups. Thus, the S728F AsBAS mutant preferentially utilises dioxidosqualene as a substrate to produce a tetracyclic triterpene when expressed in yeast. It would be interesting to introduce N731F and N731S mutations (EtBAS numbering) into the dicot BASs, and an S728N mutation into AsBAS, to determine whether these mutations alter their substrate or product specificities. However, it should be noted that *S. cerevisiae* GIL77 accumulates both 2-3-oxidosqualene and dioxidosqualene, as it lacks the ERG7 gene that encodes lanosterol synthase enzyme which converts 2,3-oxidosqualene to lanosterol (Kushiro et al., 1998). By contrast, dioxidosqualene is not detectable in yeast that are wild-type for ERG7 (e.g. BY4741 used in the present study) (Milla et al., 2002; Field and Holmlund, 1977).

AsBAS also differs from the other BASs at Y561 and N611 (EtBAS numbering), which are conserved amongst the BASs, having a Ser at each position instead (S560 and S610). Homology models and substrate docking of GsBAS1 and GsBAS2 (*Gentiana straminea*) indicated the hydroxyl of Y560 in GsBAS2 may interact with the hydroxyl of Y259 (in the MYCYCR motif) and the amido group of N610 (GsBAS2 numbering) (Liu et al., 2016). This would provide a hydrophilic environment around position *C*-13 of the oleanyl cation, which could facilitate deprotonation and thus β-amyrin formation (Liu et al., 2016). By contrast, GsBAS1 had a H560 instead which might not interact with Y269. Y560H and Y560F mutations (lacking hydroxyl groups) in GsBAS2 substantially reduced β-amyrin production, while the H560Y mutation in GsBAS1 (introducing a hydroxyl group) increased production (Liu et al., 2016). Thus, the

presence of a hydroxyl group at this position may be important for activity. AsBAS has S560 and therefore possess such a hydroxyl, but differences in steric bulk and side-chain length between Ser and Tyr may still affect activity. Furthermore, the presence of a Ser, which lacks an amido group, instead of Asn at position 610 likely also affects this interaction network. Thus, S560 and S610 are potential candidates that affect AsBAS activity.

Although there were numerous differences in amino acid sequence for AsBAS compared with the other BASs, most of the BASs had little variation in the residues known to be involved in catalysis. Importantly, protein levels will need to be determined to properly assess the differences in BAS activity. Western blots were unable to detect AsBAS, suggesting it may be expressed at very low levels or not at all, while differences in expression between the other BASs may not have been detectable using this method (Figure 4.11; see above). Thus, analysis of the protein sequences alone does not explain the observed variation in activity for the dicot BASs, but does suggest possible candidate residues for AsBAS.

4.7 Discussion

A total of twelve BASs were expressed in yeast and compared for their ability to produce β-amyrin in *S. cerevisiae*. They showed a range of β-amyrin titres, with PgBAS1 producing very small quantities of β-amyrin (titre undetermined because the peak area was below the linear region of the standard curve) and the most active homologues (AaBAS and CqBAS1) producing 10.6 mg/L β-amyrin. The lowest point of the standard curve represents a β-amyrin titre of 1.7 mg/L, so PgBAS1 produces less than this amount. The β-amyrin titres in the strains ranged at least 10-fold (between 10.6 mg/L and < 1.7 mg/L). Notably, these strains had not been metabolically engineered to enhance triterpenoid production, and expressed only the BAS from a low copy plasmid.

Interestingly, the β-amyrin titre for AaBAS was higher than that reported by Kirby *et al.*, who achieved a titre of 6 mg/L β-amyrin in *S. cerevisiae* BY4742, a haploid strain similar to BY4741 used in the present study (Kirby et al., 2008). Notably, this strain was metabolically engineered to boost triterpenoid production, overexpressing the *tHMG1* and *ERG1* genes that are rate-limiting steps in 2,3-oxidosqualene production, and repressing the *ERG7* gene which converts 2,3-oxidosqulaene to lanosterol, diverting flux from β-amyrin biosynthesis. By contrast, the non-optimised strain

expressing only *AaBAS* produced only 4 mg/L β-amyrin, ~ 2.5-fold lower than reported in the present study. Similar to the present study, Kirby *et al.* cultured cells in synthetic media supplemented with 2 % D-galactose, however the starting OD_{600} was much lower (0.02-0.03 instead of 0.2) and the culture time was longer (240 h instead of 96 h) (Kirby et al., 2008). Furthermore, they expressed *AaBAS* from the *GAL1* promoter but on a high copy plasmid instead of a low copy plasmid. These differences could explain the higher titre in the present study, as a higher copy plasmid might be expected to cause more burden, slowing cell growth, and this combined with the lower starting OD_{600} may resulted in a low cell density that limited β-amyrin production.

By contrast, Moses *et al.* and especially Dai *et al.* reported substantially higher β-amyrin titres than the current work for *MtBAS* and *GgBAS* (Moses et al., 2014b; Dai et al., 2014). Moses *et al.* expressed *MtBAS* or *GgBAS* in *S. cerevisiae* BY4742, combined with overexpression of *tHMG1* and repression of *ERG7*, and achieved β-amyrin titres of 19 mg/L or 36 mg/L, respectively (Moses et al., 2014b). By contrast, only 7.4 mg/L and 9.2 mg/L β-amyrin were produced in the *MtBAS* and *GgBAS* strains in the present study, respectively. Dai *et al.* reported a much higher β-amyrin titre of 107 mg/L in *S. cerevisiae* BY4742, using *GgBAS* and with *tHMG1*, *ERG9* and *ERG1* overexpressed (Dai et al., 2014). *ERG9* and *ERG1* are squalene synthase and squalene epoxidase and are responsible for the production squalene and is oxidation to 2,3-oxidosqulane; *ERG1* may be rate-limiting (M'Baya et al., 1989; Veen et al., 2003; Leber et al., 2001). Dai *et al.* integrated these genes and expressed them from constitutive promoters, which allowed them to grow the cells in YPD, a rich media containing glucose (Dai et al., 2014). By contrast, Moses *et al.* expressed the *BAS* and *tHMG1* from galactose-inducible promoters on a plasmid with an auxotrophic marker, requiring the use of synthetic media supplemented with galactose, which would be expected to reduce growth compared with YPD (Moses et al., 2014b). This could explain the much higher titres reported by Dai *et al.* Similar to the present work, Dai *et al.* also reported low titres of β-amyrin for *PgBAS1*, producing 1.9 mg/L β-amyrin in a strain overexpressing *tHMG1*, *ERG9* and *ERG1*, ~ 50-fold lower than for GgBAS (Dai et al., 2014).

Zhang *et al.* were also able to achieve a relatively high titre of β-amyrin (49 mg/L) with *GgBAS* (Zhang et al., 2015). To increase the supply of 2,3-oxidosqualene, they overexpressed *E. coli IDI*, *S.* cerevisiae *ERG20* and *ERG9*, and *Candida albicans*

ERG1, which together are responsible for the conversion of isopentenyl pyrophosphate (IPP), the terpene precursor produced from the MEV pathway, into 2,3-oxidosqualene. The genes were integrated and expressed from constitutive promoters engineered to contain the binding site of the *S. cerevisiae* Upc2p transcription activator, and the cultures were performed in YPD. Expression from low or high copy plasmids resulted in reduced β-amyrin titres, and plasmid loss was observed over a 48 h cultivation period (Zhang et al., 2015). Thus, integrating the pathway appears to be a promising strategy for boosting triterpenoid production.

Strain growth was also assessed, as this is an important consideration when scaling up production. The strain expressing *AsBAS* unexpectedly grew much faster than those expressing the other BASs, having a similar growth rate to the empty vector control MD-N1. Strains expressing *PgBAS1*, which produced little β-amyrin, and *BvBAS*, a multifunctional OSC that produced additional products to β-amyrin, also displayed relatively fast growth. The other grew very slowly, indicating that there was significant metabolic burden caused by BAS expression.

The growth curves of some strains (expressing *LjBAS*, *GhBAS* and *SlBAS*) showed large errors starting in exponential phase (Figure 4.10 A). This could be due to small differences in the starting OD_{600} introduced by manual error, causing some replicates to enter exponential phase earlier than others. Indeed, such differences have been shown to affect growth kinetics in microplate fermentations (Huber et al., 2009). These effects were not observed in the glucose cultures, where the strains grew much more quickly, or for the faster growing strains (*AsBAS* and MD-N1) in galactose media. It makes sense that such effects might be exacerbated in slow growing strains with longer lag phases.

BvBAS produced two additional triterpenes (U1 and U2), and U1 was detected for all BASs but PgBAS1. Analysis of the mass spectrum of U1 for BvBAS suggests that it corresponds to lupeol, and the sequence of BvBAS was consistent with this, having a Leu instead of a Trp in the MWCYCR motif. Nevertheless, to confirm the identity of U1, GC-MS analysis of BvBAS and an authentic lupeol standard must be performed. It would be interesting to perform mutational studies on BvBAS to confirm that the Leu in the MWCYCR motif is causing the shift in product specificity. In this case, a Leu to Trp mutation should shift the product specificity of BvBAS to β-amyrin.

The observed differences in productivity and growth rate could be due to variation in BAS expression levels. Increased expression would be expected to increase β-amyrin production and increase metabolic burden, reducing growth rate. Indeed, the BASs were expressed from the strong *GAL1* promoter, which could cause burden. Western blots of six BASs (AsBAS, PgBAS1, GgBAS, LjBAS, EtBAS and PtBAS) showed similar band intensities for all but AsBAS, indicating that they may be expressed at similar levels (Figure 4.11 A). AsBAS was not detected, possibly due to low expression levels or because the His-tag was inaccessible. AsBAS was the only BAS that was not codon-optimised, which could have resulted in lower expression levels than the other BASs. Low expression of AsBAS would be consistent with its minimal metabolic burden, and would indicate that it is a highly active enzyme given the relatively high β-amyrin titre (8.88 mg/L compared with 10.6 mg/L for the most productive BASs). However, this western blot did not provide quantitative information as it lacked a loading control, and so conclusions about the relative expression levels of the BASs could not be made.

To properly explain the differences in activity and investigate the causes of burden, it will be necessary to determine the relative expression levels of the BASs. Performing western blots with a suitable loading control (e.g. an anti-actin antibody) should allow for such a comparison. However, because this approach can include a significant amount of experimental error, repeated experiments would be required to obtain reliable data (Hoshino, 2017). A perhaps more accurate approach would be to perform untargeted proteomics, where proteins are extracted, digested into peptides, and analysed by LC-MS/MS. Peptides can be identified based on their mass spectra and matched to their parent protein, and the relative abundance can be measured by comparing their peak areas (label-free quantification). Such an approach could detect very low quantities of protein, possibly allowing the detection of AsBAS. Following the determination of relative expression levels, β-amyrin titres could be normalised to BAS expression and this “specific activity” used to better compare their activities.

In silico analysis of the BASs identified several amino acid variants in AsBAS that may affect its activity. First, AsBAS had a Tyr at position 533 instead of the Trp that was present in the other BASs. An EtBAS W534Y mutant was reported to have greatly reduced activity (Hoshino et al., 2017), indicating that this variant could affect AsBAS activity. Second, AsBAS had a Ser at position 728, whereas the other BASs had Asn.

An AsBAS S728F mutant showed altered substrate and product specificity when expressed in yeast (Salmon et al., 2016), and it is possible that mutating this residue to Asn would affect activity. Equally, mutants of the dicot BASs at this position, replacing Asn with Ser or Phe, might also show different activities. Finally, Y259, Y561 and N611 (EtBAS numbering) may form an interaction network that is important for activity in the dicot BAS GsBAS2 (Liu et al., 2016), and these residues are conserved across the dicot BASs tested here. AsBAS instead has S560 and S610, which would be expected to alter the interaction network and potentially catalytic activity. By contrast with AsBAS, there were no obvious amino acid variants between the other BASs that might explain the differences in activity.

Knowledge of expression levels is essential for properly interpreting the *in silico* analyses, as any differences in activity could be due to expression rather than sequence variations. It would thus be important to compare BASs with different specific activities to identify residues important for high activity. Generating homology models and performing substrate docking may provide additional information not obvious from the alignments, potentially allowing the identification of residues important for high catalytic activity in the dicot BASs. Mutational studies could then be performed to test the importance of these residues. The AsBAS amino acid variants highlighted above should also be considered with expression levels in mind and subjected to mutational studies to investigate their function. If residues that correlate with activity can be identified, this might enable protein engineering of the BASs to improve their catalytic activity.

Chapter 5: Comparison of CYP716As for oleanolic acid production

5.1 Introduction

Following the cyclisation of 2,3-oxidosqualene, the resulting triterpenes are typically subject to a range of modifications by tailoring enzymes. The most common modifications are oxygenations performed by cytochromes P450, producing triterpenoids, and glycosylations performed by UDP glycosyltransferases (UGTs), producing triterpenoid saponins. Such modifications are important for functionalising triterpenes and are responsible for the varied bioactivities observed in this family (Thimmappa et al., 2014; Moses et al., 2014a). Glycosylations occur at oxidised positions (the hydroxyl and carboxyl groups) but triterpenes such as β-amyrin have only one such group, the *C*-3 hydroxyl, which is derived from the epoxide of oxidosqualene. The P450-mediated oxygenations introduce additional moieties that can be glycosylated, increasing saponin diversity (Moses et al., 2014a). Especially important are the *C*-3 hydroxyl and *C*-28 carboxyl groups (Figure 5.1), which are the most common sites of glycosylation (Thimmappa et al., 2014).

Oleanolic acid is derived from β-amyrin through the oxidation of the *C*-28 methyl group to a carboxyl, a reaction performed by P450s of the CYP716A subfamily. This is a three-step reaction that occurs through the hydroxyl and aldehyde intermediates erythrodiol and oleanolic aldehyde (Figure 5.1). Oleanolic acid thus contains both a *C*-3 hydroxyl and a *C*-28 carboxyl group.

Like β-amyrin, oleanolic acid occurs widely in plants and is a precursor to many bioactive triterpenoids. For example, QS-21, a mixture of two saponins with a quillaic acid aglycone (containing a *C*-28 carboxyl), is being evaluated as an adjuvant in vaccines for a number of diseases, most notably in the anti-malarial vaccine RTS,S (Didierlaurent et al., 2017; Mahmoudi and Keshavarz, 2017). It is also used in a commercially available veterinary vaccine against feline leukaemia virus (Sun et al., 2009). Furthermore, several oleanolic acid derivatives are under investigation for various indications, and some have undergone clinical trials (Moses et al., 2013). For example, the synthetic derivative bardoxolone methyl has been investigated in clinical trials as an anti-cancer agent and for the treatment of chronic kidney disease (Moses et al., 2013). Meanwhile, the synthetic derivative RTA 408 (omaveloxolone) is

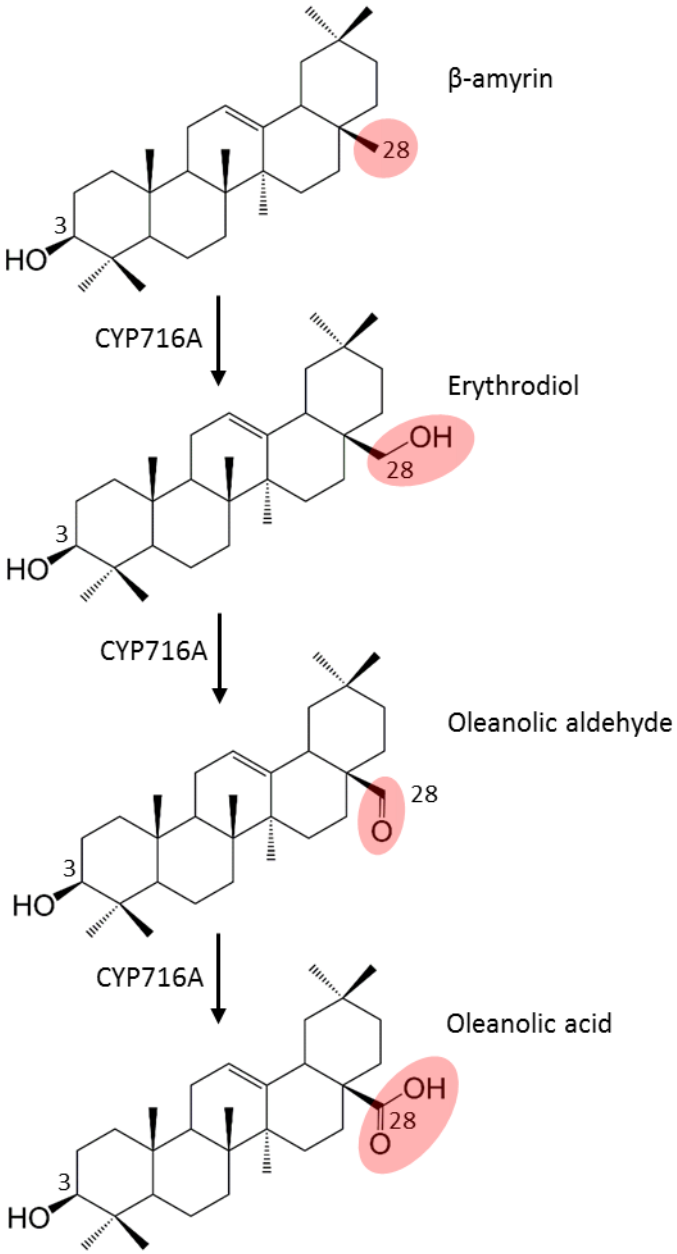


Figure 5.1. Oxidation of β-amyrin to oleanolic acid by cytochromes P450 of the CYP716A subfamily. Oxidation takes place at the C-28 methyl group and occurs through hydroxyl and aldehyde intermediates (erythrodiol and oleanolic aldehyde) to generate a carboxyl group (oleanolic acid). Carbon 3, which contains the hydroxyl group derived from 2,3-oxidosqualene, and carbon 28 are labelled.

currently undergoing a phase II clinical trial for the treatment of the neurodegenerative disease Friedreich's ataxia, and has shown promising results (ClinicalTrials.gov, Identifier NCT02255435, 2014; Strawser et al., 2017). Saponins derived from oleanolic acid and hederagenin (an oleanolic acid derivative) have been reported as having anti-feedant and insecticidal activities (Shinoda *et al.*, 2002; Nielsen *et al.*, 2010; Augustin *et al.*, 2012), potentially lending themselves to agricultural applications. Oleanolic acid itself is reported to have hepatoprotective properties and is used in China to treat liver disorders such as viral hepatitis (Pollier and Goossens, 2012). For more information, see section 1.2 and Figure 1.6.

The CYP716A subfamily comprises many characterised P450s involved in triterpenoid biosynthesis (Miettinen et al., 2017). Most catalyse the *C*-28 oxidation of β-amyrin and related triterpenes (often α-amyrin and lupeol), however, some catalyse other reactions (Miettinen et al., 2017; Ghosh, 2017). For example, CYP716A2 from *Arabidopsis thaliana* catalyses the *C*-22α hydroxylation of β-amyrin (Yasumoto et al., 2016), while CYP716A14v2 (*Artemisia annua*) oxidises the *C*-3 hydroxyl group of β-amyrin and α-amyrin to a ketone (Moses et al., 2015a). CYP716A141 from *Playocodon grandiflorus* catalyses the *C*-16β hydroxylation of β-amyrin and oleanolic acid, but also has a weaker *C*-28 oxidation activity against β-amyrin (Tamura et al., 2017; Miettinen et al., 2017). Nevertheless, the majority of CYP716As discovered to date are *C*-28 oxidases.

Due to the importance of the *C*-28 oxidation reaction in triterpenoid biosynthesis, I sought to compare members of this subfamily for their ability to oxidise β-amyrin to oleanolic acid when expressed in yeast. Although a large number of CYP716As have been characterised, many were discovered relatively recently and metabolic engineering studies have typically utilised the same subfamily member (CYP716A12 from *Medicago truncatula*) for the production of oleanolic acid and its derivatives (Moses et al., 2014b; Dai et al., 2014; Fukushima et al., 2013; Arendt et al., 2017). However, as shown earlier for the comparative analysis of β-amyrin synthases (see Chapter 4), the choice of homologue can have a large impact on strain productivity. This is especially true of the CYP716As, which also produce the intermediate compounds erythrodiol and oleanolic aldehyde, meaning that not only catalytic activity but also product ratios might differ between homologues.

Sixteen CYP716As from a range of plants were compared for their ability to oxidise β-amyrin to oleanolic acid when expressed in yeast. They were co-expressed with AaBAS, the most productive BAS identified in the BAS comparative analysis (see Chapter 4), and the cytochrome P450 reductase ATR2 from *Arabidopsis thaliana*. The CYP716As spanned the majority of identified CYP716As, including the commonly used CYP716A12. In addition, two new CYP716As, from *Theobroma cacao* (cocoa tree) and *Olea europaea* (olive), were functionally characterised by expression in yeast and included in the comparative analysis. Triterpenoid production was monitored by GC-MS and microplate growth assays were performed to assess burden.

5.2 Characterisation of two new cytochromes P450 of the CYP716A subfamily

An initial investigation was untaken to identify additional *C*-28 oxidises from the CYP716A subfamily. A BLASTp search on NCBI against the non-redundant database with default parameters was performed using CYP716A12, a characterised β-amyrin *C*-28 oxidise, as the query sequence. This returned 100 hits with sequence identities to CYP716A12 > 70 %, many of which were annotated as predicted β-amyrin *C*-28 oxidases, and two genes were selected for analysis. The first was CYP716A48 from *Olea europaea* (olive; BAP59949), which had 78 % amino acid identity to CYP716A12. *O. europaea* is known to produce oleanolic acid and is used as a source of this triterpenoid for commercial purposes (Pollier and Goossens, 2012). It encodes a multifunctional triterpene synthase that produces mainly α-amyrin and β-amyrin (Saimaru et al., 2007) (no monofunctional BAS has been reported), but the gene responsible for oleanolic acid biosynthesis was yet to be identified. The high sequence identity of CYP716A48 with characterised *C*-28 oxidases (> 70 %) (Figure 5.4), and the known oleanolic acid production capacity of *O. europaea*, suggest that CYP716A48 may indeed be a β-amyrin *C*-28 oxidase. Furthermore, it contains the expected conserved residues for a P450, notably the proton transfer groove, ExxR motif, PERF motif and haem binding loop, as well as a predicted N-terminal transmembrane anchor (Figure 5.4).

A predicted β-amyrin *C*-28 oxidase from *Theobroma cacao* (cocoa tree; XP_017978138.1) with 74 % amino acid identity to CYP716A12 was located in close proximity with another predicted β-amyrin *C*-28 oxidase (CYP716A147;

XP_007023618), the putative β-amyrin synthase identified earlier (TcBAS; see Chapter 4), a predicted P450 of the CYP705A subfamily, and a predicted cycloartenol synthase on the *T. cacao* genome, in what appeared to be a possible gene cluster (within a 270 kb region). Cycloartenol synthases cyclise oxidosqualene to the plant sterol cycloartenol, and triterpene synthases such as BASs are believed to have evolved from cycloartenol synthases via gene duplication (Augustin et al., 2011). Gene clusters have been described for triterpenoid and other plant secondary metabolic pathways before (Qi et al., 2004; Field and Osbourn, 2008; Field et al., 2011; Osbourn and Field, 2009), and contain genes involved in the same biosynthetic pathway. The presence of this putative gene cluster suggests that these predicted enzymes may be genuine. CYP716A147 had been simultaneously identified (but not characterised) by Miettinen *et al.* (Miettinen et al., 2017) and was selected for analysis. It had a high amino acid identity (> 70 %) to CYP716A12 and other characterised CYP716As, and contained the expected conserved residues that were also observed for CYP716A48 (Figure 5.4, Figure 5.5).

The *CYP716A48* and *CYP816A147* genes were synthesised with codon-optimisation for *S. cerevisiae* (see materials and methods), and were cloned into a centromeric plasmid with the *Euphorbia tirucalli* BAS (*EtBAS*) and *ATR2*, a cytochrome P450 reductase (CPR) from *Arabidopsis thaliana* that is required for P450 activity. Each gene was expressed from a galactose-inducible promoter. Most plant P450s, including those involved in triterpenoid biosynthesis, require a plant CPR for activity in yeast. The CPR oxidises NADPH and transfers the electrons to the P450 active site, where they are used to activate molecular oxygen to oxidise the substrate (Werck-Reichhart and Feyereisen, 2000; Jensen and Møller, 2010; Bak et al., 2011). While yeast encodes its own native CPR (*NCP1*), it is less effective at transferring electrons to heterologously expressed plant P450s, resulting in lower catalytic activity (Dejong et al., 2005). Instead, a plant CPR is often co-expressed with the P450, and the *A. thaliana* CPRs *ATR1* and *ATR2* are commonly used for this purpose (Urban et al., 1997; Zhao et al., 2016; Dai et al., 2013, 2014; Moses et al., 2014b; Arendt et al., 2017; Gavira et al., 2013; Wang et al., 2015); ATR2 was reported to have a higher catalytic activity than ATR1 (Schückel et al., 2012; Zhu et al., 2018).

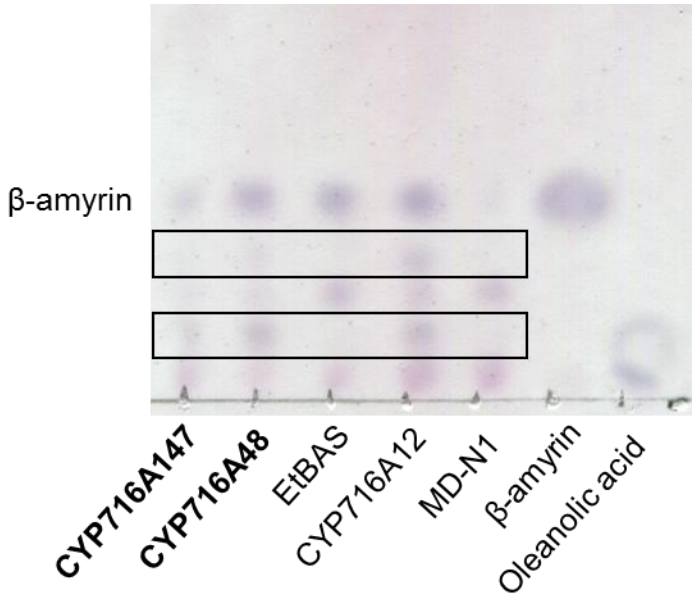


Figure 5.2. Thin layer chromatography (TLC) indicates that CYP716A147 and CYP716A48 are functional in yeast. TLC was performed on extracts of cells expressing oleanolic acid pathways containing either *CYP716A147*, *CYP716A48* or *CYP716A12* (a previously characterised P450). A strain expressing *EtBAS* only, the control strain MD-N1, and 0.6 µg of β-amyrin and oleanolic acid standards were also analysed. The boxes indicate additional bands present in the oleanolic acid strains that were not present for *EtBAS* and the control strain. These could represent erythrodiol, oleanolic aldehyde and/or oleanolic acid.

The plasmids expressing *GgBAS*, P450 (*CYP716A147* or *CYP716A48*) and *ATR2* were transformed into *S. cerevisiae* BY4741, and the resulting strains (MD-OA5 and MD-OA6) were cultured and prepared for TLC analysis as before (see Chapter 4). Strains expressing *EtBAS* (MD-BA5), an oleanolic acid pathway containing *CYP716A12*, *GgBAS* and *ATR2* (MD-OA1), and an empty vector control (MD-N1) were run as controls. Extracts of each strain, as well as authentic β-amyrin and oleanolic acid standards, were loaded onto the TLC plate. As expected, EtBAS and MD-N1 had multiple bands, one of which was much stronger for EtBAS and migrated a similar distance to the β-amyrin standard (Figure 5.2). The CYP716A12 strain had two additional bands that migrated slower than β-amyrin, and likely corresponded to products of the β-amyrin *C*-28 oxidation reaction (erythrodiol, oleanolic aldehyde, or oleanolic acid). Accordingly, the oleanolic acid standard migrated more slowly than the β-amyrin standard. Unfortunately, this standard gave a very diffuse band, possibly

due to manual error when loading it onto the plate, making it difficult to identify a putative oleanolic acid band in the samples. Both CYP716A147 and CYP716A48 gave a similar pattern of bands to CYP716A12, containing the putative β-amyrin band as well as those possibly corresponding to β-amyrin derivatives (Figure 5.2). The data thus indicate that both P450s are active in yeast.

To confirm the TLC results and identify the products of CYP716A147 and CYP716A48, these strains were analysed by GC-MS. Cells were cultured and GC-MS analysis was performed as before (see Chapter 4). The MD-N1 control strain was also analysed, as was a standards mixture containing β-amyrin, erythrodiol, oleanolic acid, ergosterol and coprostanol (an oleanolic aldehyde standard is not available). Peaks with the same retention times as the erythrodiol and oleanolic acid standards were observed for both CYP716A147 and CYP716A48 (Figure 5.3 A), and their identities were confirmed by comparison of their mass spectra with the authentic standards (Figure 5.3 B-E). An additional peak that was not present in MD-N1 was also observed (Figure 5.3 A), and its mass spectrum was consistent with that reported for an authentic oleanolic aldehyde standard (Figure 5.3 F-G and see Fig. S3 of (Khakimov et al., 2015)). A peak that eluted shortly after oleanolic aldehyde was also observed (Figure 5.3 A), but this was present in MD-N1 and probably corresponds to a native yeast metabolite. Thus, CYP716A147 and CYP716A48 are genuine β-amyrin *C*-28 oxidases, oxidising β-amyrin to produce oleanolic acid via erythrodiol and oleanolic aldehyde.

Figure 5.3. GC-MS confirms that CYP716A147 and CYP716A48 are functional in yeast. **(A)** Total ion chromatograms of strains expressing *CYP716A147* and *CYP716A48*, the control strain MD-N1, and a mixture of ergosterol, β-amyrin, erythrodiol and oleanolic acid standards (each at 50 µg/ml). Peaks corresponding to erythrodiol and oleanolic acid were observed for both P450s, as was an additional peak not present in the control strain that may correspond to oleanolic aldehyde. **(B-E)** Mass spectra of the erythrodiol and oleanolic acid peaks from the CYP716A strains, showing that they match the mass spectra of the standards. **(F-G)** Mass spectra of the putative oleanolic aldehyde peaks. These are consistent with what would be expected for oleanolic aldehyde (e.g. see Fig. S3 of Khakimov *et al.* for the mass spectrum of an authentic oleanolic aldehyde standard (Khakimov et al., 2015)).

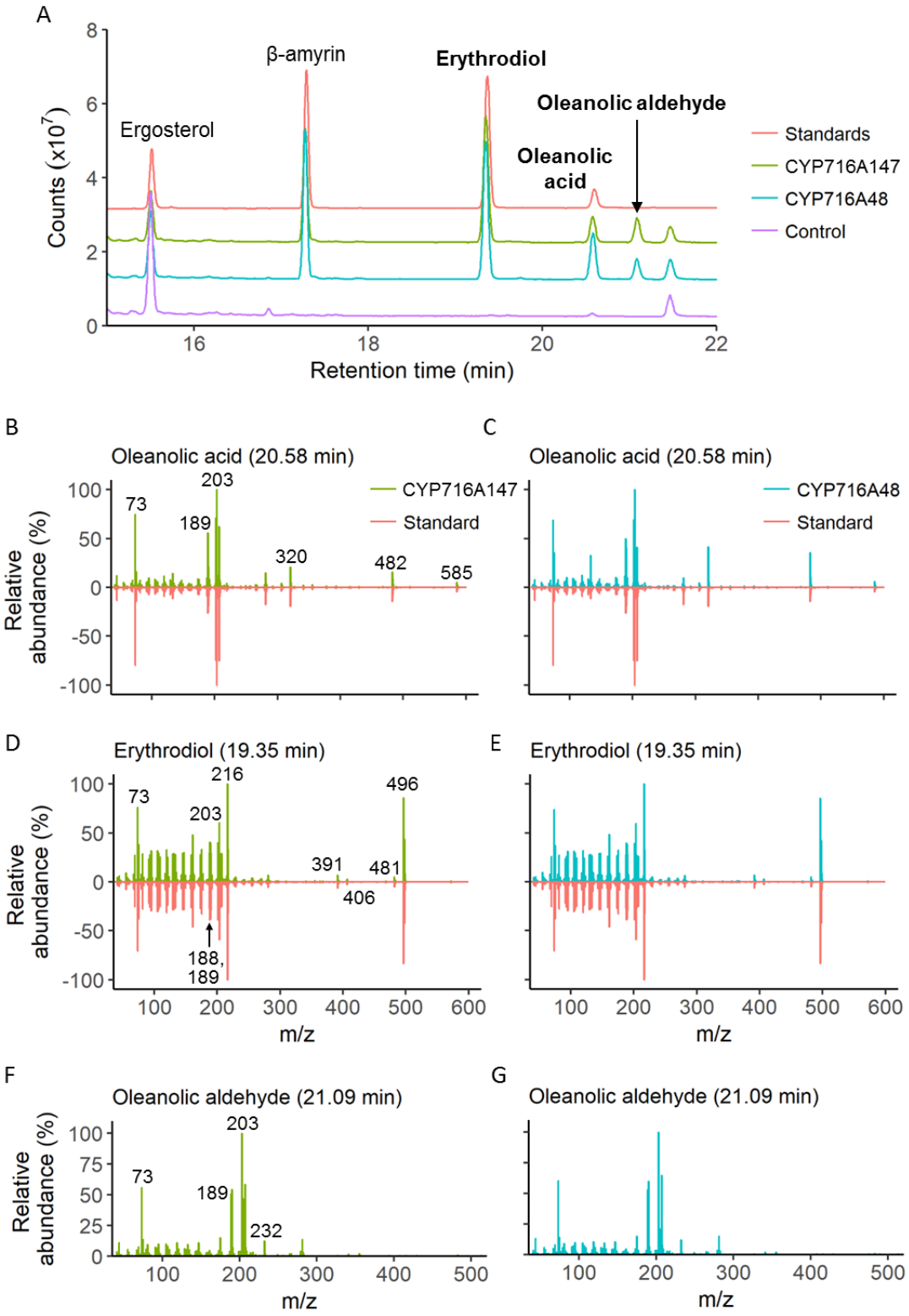


Figure 5.3. GC-MS confirms that CYP716A147 and CYP716A48 are functional in yeast.

See previous page for full legend

5.3 Comparison of CYP716A homologues for the production of oleanolic acid

Sixteen CYP716As (Table 5.1) were compared in yeast for their ability to oxidise β-amyrin to oleanolic acid. These represented many of the CYP716As characterised to date, including CYP716A12 and the newly discovered CYP716A147 and CYP716A48. A multiple sequence alignment of the protein sequences was performed (Figure 5.4), and most had sequence identities of 65 % or greater (Figure 5.5). The exceptions to this were CYP716A1 and CYP716A80, which had lower sequence identities to the other CYP716As but an 80 % sequence identity to each other. Phylogenetic analysis of the CYP716As and numerous other P450s involved in triterpenoid biosynthesis showed that the CYP716As clustered together (Figure 5.6). CYP716A1 and CYP716A80 formed a separate clade with CYP716A2, which has *C*-16α, *C*-22α and *C*-28 hydroxylation activity, while CYP716A110 was slightly more distantly related.

The CYP716As were synthesised with codon-optimisation for *S. cerevisiae* and cloned into centromeric plasmids together with *AaBAS* and *ATR2*; AaBAS was the most productive BAS identified in Chapter 4 (together with CqBAS1). *AaBAS* and the CYP716As were expressed from the strong *pGAL1* and *pGAL10* galactose-inducible promoters, while ATR2 was expressed from *pGAL3*, which was reported to be weaker (Paddon et al., 2013). Previous studies have indicated that the ratio of P450 to CPR is important for activity. Moses *et al.* compared different *ATR1* expression levels for triterpenoid production, integrating it into the chromosome and expressing it from low and high copy plasmids, while the P450 (*CYP716Y1*) was expressed from a high copy plasmid. The strain expressing *ATR1* from the low copy plasmid gave the highest titres (Moses et al., 2014b). Paddon *et al.* found that reducing CPR expression improved the growth and viability of an artemisinic acid production strain, but also reduced titres (Paddon et al., 2013). Indeed, when the CPR is present in excess over the P450 uncoupling in electron transfer can occur, and the resulting electron leakage generates reactive oxygen species that are toxic to the cells (Renault et al., 2014). Consistent with this, studies of natural systems have shown the P450 to be present in greater abundance than the CPR (Peterson et al., 1976; Reed and Backes, 2012; Renault et al., 2014).

Table 5.1. CYP716A homologues tested in this study. The cytochrome P450 reductase ATR2 is also shown.

CYP716A	Species	Accession number	Reference
CYP716A12	*Medicago truncatula*	CBN88269.1	Fukushima (2011)
CYP716A15	*Vitis vinifera*	BAJ84106	Fukushima (2011)
CYP716AL1	*Catharanthus roseus*	AEX07773	Huang (2012)
CYP716A52v2	*Panax ginseng*	AFO63032	Han (2012, 2013)
CYP716A147	*Theobroma cacao*	XP_007023618	The present study
CYP716A48	*Olea europaea*	BAP59949.1	The present study
CYP716A44	*Solanum lycopersicum*	AK329870.1	Yasumoto (2017)
CYP716A75	*Maesa lanceolata*	AHF22088.1	Moses (2015a)
CYP716A79	*Chenopodium quinoa*	ANY30854.1	Fiallos-Jurado (2016)
CYP716A80	*Barbarea vulgaris*	ALR73782.1	Khakimov (2015)
CYP716A83	*Centella asiatica*	AOG74832.1	Miettinen (2017), Kim (2018)
CYP716A110	*Aquilegia coerulea*	AOG74847.1	Miettinen (2017)
CYP716A140	*Platycodon grandiflorus*	AOG74836.1	Miettinen (2017)
CYP716A179	*Glycyrrhiza uralensis*	BAW34647.1	Tamura (2017b)
CYP716A244	*Eleutherococcus senticosus*	APZ88353.1	Jo (2017)
CYP716A1	*Arabidopsis thaliana*	NP_198460.1	Yasumoto (2016)
ATR2	*Arabidopsis thaliana*	NP_849472.2	Urban (1997)

```
CYP716A80       ----MYLT-ILFLFVSSILLSLMFLLRKHLSH------FSYQNLPPGKTGFPLIGESLSF     49
CYP716A1        ----MYMAIMIILFLSSILLSLLLLLRKHLSH------FSYPNLPPGNTGLPLIGESFSF     50
CYP716A110      -MEQISLSWLITLVALFLFLPLILLSTRNK-------ANTSNNFPPGKTGWPIIGESLEF     52
CYP716A79       -MELFFLCSIA--LVFFLSLPFLYLFYRHNS-------TRGYKLPPGSMGWPVVGESLEF     50
CYP716A12       MEPNFYLSLLLLFVSF-ISLSLFFIFYK-Q--------KSPLNLPPGKMGYPIIGESLEF     50
CYP716A179      -MEHFYMSLLLLFVTL-VSLSLFFLIFYYNKH----NMNNNNNLPPGKMGYPVIGESLEF     54
CYP716A147      -MDPIIL-LFC--TIFFAAISLGVLLYSKNP------NASHPNLPPGRMGLPLIGESLEY     50
CYP716A75       -MELLFVSLLSLFLLILLPLSLLFLFPSSFSTTTTEANKNSANLPPGLTGWPVVGESFQF     59
CYP716A140      -MELLYVSLLTLLVLL-FPLSLHFLFKKTKS-----GLGEVGRLPPGKTGWPIVGESIEF     53
CYP716A15       -MEVFFLSLLLIFVLS-VSIGLHLLFYKHRS-----HFTG-PNLPPGKIGWPMVGESLEF     52
CYP716A83       -MELFFVPLFSSLVLF-VFFCFLLLFYKNNK-----WRSSGAPLPPGQTGWPFIGESYEF     53
CYP716A52v2     -MELFYVPLLSLFVLF-ISLSFHFLFYKSKP-----SSSGGFPLPPGKTGWPIIGESYEF     53
CYP716A244      -MQLFYVPLLSLFVLL-VSLSFYFLFYKSKS-----GSPGALPLPPGKTGWPVIGESFEF     53
CYP716A44       -MELLYVCLVCVFVFL-VSLLL---LYKKK-------SGE--GLPPGKTGWPVFGESLEF     46
CYP716AL1       -MEIFYVTLLSLFVLL-VSLSFHFLFYKNKS-----TLPG--PLPPGRTGWPMVGESLQF     51
CYP716A48       -MEFFYVSLLCLFVFL-ISLSLHFLFYKNKS-----SFSG--QIPPGKTGWPVIGESLEF     51
                    : :  .         : :                    :***  * *..*** .:

CYP716A80       LSEGSQGHPEKFITDRVRRFISSSSGVFKTHLFGSPTAVMTGASGNKFLFTNENKLVVSW     109
CYP716A1        LSAGRQGHPEKFITDRVRRFSSSSSCVFKTHLFGSPTAVVTGASGNKFLFTNENKLVVSW     110
CYP716A110      LNTGRRGVPEKFIHDRMQKFS---SHIFRTSLLGEPAAVLCGPTGNKFLYSNENKLVQAW     109
CYP716A79       FSTGWKGYPEKFIFDRLKKYKP--SQVFKTSIFGEQVAILCGATGNKFLYSNENKLVQAW     108
CYP716A12       LSTGWKGHPEKFIFDRMRKYS---SELFKTSIVGESTVVCCGAASNKFLFSNENKLVTAW     107
CYP716A179      LSTGWKGHPEKFIFDRMVRYS---SELIKTSILGVPTVIFCGPACNKFLFSNENKLVTGW     111
CYP716A147      LLTGRKGYPEKFLNDRMAKYS---SQVFKTSILGESMAVVCGAAGNKFLFSNENKLVTAW     107
CYP716A75       LAAGWRGNPEKFIFDRIAKYS---SYVYKTNLMLERTAVFCGAPAHKFLFSNENKLVQSW     116
CYP716A140      LSTGWKGHPEKFIFDRMSKYS---PQVFRTSLMLEDAAVFCGSAGNKFLFSNEKQLVTAW     110
CYP716A15       LSTGWKGHPEKFIFDRISKYS---SEVFKTSLLGEPAAVFAGAAGNKFLFSNENKLVHAW     109
CYP716A83       LSTGWKGYPEKFIFDRIAKHS---SNVFKTSILGEHAAVFCGAACNKFLFSNENKLVQAW     110
CYP716A52v2     LSTGWKGYPEKFIFDRMTKYS---SNVFKTSIFGEPAAVFCGAACNKFLFSNENKLVQAW     110
CYP716A244      LSTGWKGHPEKFIFDRLAKYS---SNVFRTSLLGQPAAVFCGAACNKFLFSNENKLVQAW     110
CYP716A44       LSCGWKGHPEKFIFDRVAKYS---SSVFKTHLLGEEAAVFCGASANKFLFSNENKLVQAW     103
CYP716AL1       LSAGWKGHPEKFIFDRMAKYS---SNVFRSHLLGEPAAVFCGAIGNKFLFSNENKLVQAW     108
CYP716A48       LSNGWKGHPEKFIFDRIAKYS---SYVFRTHLFGEPAAVFCGANGNKFLFSNENKLVQAW     108
                :  * :* ****: **: :.      : :: :.    .:  *   :***::**::** .*

CYP716A80       WPDSVKKIFPYTQ-ST-YTEESKKLRILLLQFMKPEALRKYIGVMDEVTQRHFETEWTNK     167
CYP716A1        WPDSVNKIFPSSM-QTSSKEEARKLRMLLSQFMKPEALRRYVGVMDEIAQRHFETEWANQ     169
CYP716A110      WPSSVYKIFPSTG-QSSSSEEAIKMRKMLLSFFKPEALQRYVETMDTIAKQHLKNEWENR     168
CYP716A79       WPKSVDKIFPAAT-QHSSIEEARTMRKLIPLFLKPESLQRYIPIMDTIATRHMESGWEGK     167
CYP716A12       WPDSVNKIFPTTSLDSNLKEESIKMRKLLPQFFKPEALQRYVGVMDVIAQRHFVTHWDNK     167
CYP716A179      WPDSVNKIFPTTS---NSKEESKKMRKLLPQFLKPEALQRYVGIMDTLAQKHFASLWEEK     168
CYP716A147      WPDSVNKIFPSST-QTSSKEESKKMRKMLPNFLKPEALQRYIGMMDNIAQRHFEASWEGK     166
CYP716A75       WPSSVNKIFPSSN-QTSSKEEAMKMRKMLPNFFKPEALQGYIGIMDTIAQRHFAADWDNK     175
CYP716A140      WPASVDKVFPSSD-QR-SKEEAIKMKKLLPTFFKPVALQNYVGVMDQITERHFASDWDNN     168
CYP716A15       WPSSVDKVFPSST-QTSSKEEAKKMRKLLPQFFKPEALQRYIGIMDHIAQRHFADSWDNR     168
CYP716A83       WPDSVNKVFPSST-QTSSKEEAIKMRKMLPNFLKPDALQRYVGTMDSIARRHFESGWDNK     169
CYP716A52v2     WPDSVNKVFPSST-QTSSKEEAIKMRKMLPNFFKPEALQRYIGLMDQIAANHFESGWENK     169
CYP716A244      WPDSVNKVFPSST-QTSSKEEAIKMRKLLPNFLKPEALQRYVGIMDQIAKKHFESGWENK     169
CYP716A44       WPNSVNKVFPSST-QTSSKEEAIKMRKMLPNFFKPEALQRYVGIMDHITQRHFASGWENK     162
CYP716AL1       WPDSVNKVFPSSN-QTSSKEEAIKMRKMLPNFLKPEALQRYIGLMDQIAQKHFSSGWENR     167
CYP716A48       WPASVDKVFPSSN-QTSSKEEAVKMRKMLPTFFKPEALQRYVGIMDHIAQRHFSDGWDNK     167
                ** ** *:** :       **: .:: ::  *:** :*: *:  ** :: .*:   *  .
```

```
CYP716A80       KQLIVYPLSKKLTFSIACRLFLSMDDPERVSKLEERFNSVVMGIYSIPIDLPGTRFNRSI   227
CYP716A1        DQVIVFPLTKKFTFSIACRSFLSMEDPARVRQLEEQFNTVAVGIFSIPIDLPGTRFNRAI   229
CYP716A110      KEVMVFPLSKGYTFALACKLFMSIDDPIQLARFADPFAILASGVIAIPINLPGTSFNKAI   228
CYP716A79       DKVEVFPLAKRYTFWVACRLFLSIEDPVHVAKFADPFNEIAAGIISIPIDLPGTPFHKGI   227
CYP716A12       NEITVYPLAKRYTFLLACRLFMSVEDENHVAKFSDPFQLIAAGIISLPIDLPGTPFNKAI   227
CYP716A179      THVTVYPLAKRYTFMLACRLFMSVEDENHVAKFREPFHLLASGIISVPIDLPWTPFNRGI   228
CYP716A147      QEITVFPLAKRYTFWVACKVFLSIEDPEHVSKFADPFNALASGIISVPINLPGTPFRRAI   226
CYP716A75       DYIVVFPLCKRYTFWLACKIFMSIEDPKDVDRFLARFNLVAEGLLSIPIDLPGTPFHHSI   235
CYP716A140      NK-VVYPLTKRFTFALACKIFVSIDDPDEVARLVVPFESIATAILSIPLDLPGTGFRRGI   227
CYP716A15       DEVIVFPLAKRFTFWLACRLFMSIEDPAHVAKFEKPFHVLASGLITVPIDLPGTPFHRAI   228
CYP716A83       NEIVVFPLAKTYTFWIACKLFVSVEEPSQVAKLLEPFSAIASGIISVPIDLPGTPFNSAI   229
CYP716A52v2     NEVVVFPLAKSYTFWIACKVFVSVEEPAQVAELLEPFSAIASGIISVPIDLPGTPFNSAI   229
CYP716A244      KEVSVFPLAKNYTFWIACKVFLSVEEPTQVAKLLEPFNAIASGIISVPIDLPGTPFNSAI   229
CYP716A44       EQVVVFPLTKRYTFWLACRLFLSVEDPKHVAKFADPFDVLASGLISIPIDLPGTPFNRAI   222
CYP716AL1       EQVEVFPLAKNYTFWLASRLFVSVEDPIEVAKLLEPFNVLASGLISVPIDLPGTPFNRAI   227
CYP716A48       NEVVVFPLAKRYTFWLACRLFVSVEDPAHVAKFADPFNELASGLISIPIDLPGTPFHRAI   227
                    *:** *  ** :*.: *:*:::   : .:   *  :. .: ::*::** * *. .*

CYP716A80       KASRLIRKEVCAIIGQRREELKAGRASAEQDVLSHMLTSVGE-----TKDEDLANYLIGI   282
CYP716A1        KASRLLRKEVSAIVRQRKEELKAGKALEEHDILSHMLMNIGE-----TKDEDLADKIIGL   284
CYP716A110      KASELIRKDLHKIIKQRKIELAEKRASPMQDVLSHMLSTS-DENGTFMSEMDIADKILGL   287
CYP716A79       KSSEIVRKELRAIIKQRKLDFADGKASPTQDILSHMLLTS-TEDGKFMNEMDIANKILGL   286
CYP716A12       KASNFIRKELIKIIKQRRIDLAEGTASPTQDILSHMLLTS-DENGKSMNELNIADKILGL   286
CYP716A179      KASNFIRKELLKIIRQRKVDLAQGVASPTQDILSHMLLTCDDENGEFMTELNIADKILGL   288
CYP716A147      NASELIRKELMAIIKQRKIDLAENKAAPNQDILSHMLLAT-DENGQYLNELNIADRILGL   285
CYP716A75       KAAEFIREHLVAIIKQRKIDLAEGKASPTQDIMSYMLLTP-DEDGKFMKEYDIADKILGL   294
CYP716A140      NAANFIRKELLAIIKQRKIDLAKGKASPTQDILSHMLLTS-DENGKFMQEYDIADKILGL   286
CYP716A15       KASNFIRKELRAIIKQRKIDLAEGKASQNQDILSHMLLAT-DEDGCHMNEMEIADKILGL   287
CYP716A83       KSSKRIREMLWKIIKQRKIDLAEGKASPTQDILSHMLLTS-DENGKFMSELDIADKILGL   288
CYP716A52v2     KSSKIVRRKLVGIIKQRKIDLGEGKASATQDILSHMLLTS-DESGKFMGEGDIADKILGL   288
CYP716A244      KSSKIIRDKLLGIIKQRKIDLGEGKASPTQDILSHMLLTS-DENGKFMTEGDIADKILGL   288
CYP716A44       KASNFIRKELVRIIKQRKIDLGEGKVSSTQDILSHMLLTC-DENGKFLGDLDIADKILGL   281
CYP716AL1       KASNQVRKMLISIIKQRKIDLAEGKASPTQDILSHMLLTS-DENGKFMHELDIADKILGL   286
CYP716A48       KSSNFIRKELVSIIKQRKIDLAEGKASPTQDILSHMLLTS-DESGKFMHELDIADKILGL   286
                :::. :*  :  *: **: ::    .   :*::*:**           : ::*: ::*:

CYP716A80       LIGGHDTAAIATTFIISYLAEYPHVYQRVLQEQKEILNEKKEKERLKWEDIEKMKYSWNV   342
CYP716A1        LIGGHDTASIVCTFVVNYLAEFPHVYQRVLQEQKEILKEKKEKEGLRWEDIEKMRYSWNV   344
CYP716A110      LIGGHDTASAAITFVMKFLAELPHIYNDVHKEQMEILKSKGSRELLKWEDIQKMKYSWNV   347
CYP716A79       LIGGHDTASASCTFIVKFLAELPHIYEGVYKEQMEIANSKKPGELLNWEDIQKMKYSWNV   346
CYP716A12       LIGGHDTASVACTFLVKYLGELPHIYDKVYQEQMEIAKSKPAGELLNWDDLKKMKYSWNV   346
CYP716A179      LIGGHDTASAACTFIVKYLAELPHIYDRVYQEQMEIANSKSPGELLNWDDINKMRYSWNV   348
CYP716A147      LIGGHDTASAAITFIIKYLAELPDIYNEVYKEQMEIARSKEPGELLNWEDIQKMKYSWNV   345
CYP716A75       LVGGHDTASSACAFIVKYLAELPQVYQGVYKEQMEIAKSKGPGELLNWDDIQKMKYSWNV   354
CYP716A140      LIGGHDTASSACVFIVKYLAELPEIYDGVYKEQMEIAESKAPGELLNWDDLTKMKYSWNV   346
CYP716A15       LIGGHDTASAAITFLIKYMAELPHIYEKVYEEQMEIANSKAPGELLNWDDVQNMRYSWNV   347
CYP716A83       LIGGHDTASSACTFVVKFLAELPEIYDGVYKEQMEIVKSKGPGELLNWDDIQKMKYSWNV   348
CYP716A52v2     LIGGHDTASSACTFVVKFLAELPQIYEGVYQEQMEIVKSKKAGELLKWEDIQKMKYSWNV   348
CYP716A244      LIGGHDTASSACTFVVKFLAELPEIYEGVYKEQMEIAKSKKPGELLSWEDIQKMKYSWNV   348
CYP716A44       LIGGHDTASSACSFIVKYLAELPHIYQRVYTEQMEIAKSKGPGELLRWEDIQKMKYSWNV   341
CYP716AL1       LIGGHDTASSACTFIVKFLGELPEIYEGVYKEQMEIANSKAPGEFLNWEDIQKMKYSWNV   346
CYP716A48       LVGGHDTASSACTFVVKYLAELPEIYEGVYQEQMEIAKSKAPGELLNWDDIQKMKYSWNV   346
                *:******:    *::.::.* *.:*: *  ** ** ..*   * * *:*: :*:*****
```

```
CYP716A80       ACEVMRLVPPLTGTFREAIDHFTFKGFYIPKGWKLYWSATATHMNPDYFPEPERFEPNRF    402
CYP716A1        ACEVMRIVPPLSGTFREAIDHFSFKGFYIPKGWKLYWSATATHMNPDYFPEPERFEPNRF    404
CYP716A110      ACEVMRLAPPLQGGFREALADFTYNGFSIPKGWKLYWSTNSTHKNSTYFPNPENFDPSRY    407
CYP716A79       ACEVMRLAPPLQGGFREAISDFMYNGFQIPKGWKLYWSANTTHLNPECFPEPQKFDPSRF    406
CYP716A12       ACEVMRLSPPLQGGFREAITDFMFNGFSIPKGWKLYWSANSTHKNAECFPMPEKFDPTRF    406
CYP716A179      ASEVMRVAPPLQGGFREAINDFVFNGFSIPKGWKLYWSANSTHKNPEYFPAPEKFDPTRF    408
CYP716A147      ACEVMRLAPPLQGAFREAITDFTFSGFSIPKGWKLHWNVNSTHKNVECFPEPEKFDPTRF    405
CYP716A75       ACEVLRLAPPLQGAFRDVLKDFMYEGFYIPKGWKVYWSAHSTHKNPEYFPEPYKFDPSRF    414
CYP716A140      ACEVLRLAPPLQGAFREAITDFMYSGFSIPKGWKLYWSANSTHKNPKFFPEPQKFDPSRF    406
CYP716A15       ACEVMRLAPPLQGAFREAITDFVFNGFSIPKGWKLYWSANSTHKSPECFPQPENFDPTRF    407
CYP716A83       ACEVLRLAPPLQGGFREVLTDFSYNGFSIPKGWKIYWTANSTHRNSEVFPEPLKFDPSRF    408
CYP716A52v2     ACEVLRLAPPLQGAFREALSDFTYNGFSIPKGWKLYWSANSTHINSEVFPEPLKFDPSRF    408
CYP716A244      ACEVLRLAPPLQGAFREVLTDFSYNGFSIPKGWKLYWSANSTHRNSEVFPEPLKFDPSRF    408
CYP716A44       ACEVLRLAPPLQGAFREALSDFMFNGFYIPKGWKIYWSANSTHKREEFFPDPEKFDPSRF    401
CYP716AL1       ACEVLRLAPPLQGAFREALNDFMFHGFSIPKGWKIYWSVNSTHRNPECFPDPLKFDPSRF    406
CYP716A48       ACEVLRLAPPLQGAFREAITDFMFNGFSIPKGWKLYWSANSTHRNSEFFPEPLKFDPSRF    406
                *.**:*: *** * **:.: .* : ** ******::*.. :**     ** * .*:*.*:

CYP716A80       EGSGPKPYTYIPFGGGPRMCPGREFARLEILVIIHNLVNRFKWEKVSPNENKIVVDPLPK    462
CYP716A1        EGSGPKPYTYVPFGGGPRMCPGKEYARLEILIFMHNLVNRFKWEKVFPNENKIVVDPLPI    464
CYP716A110      EGNGPAPYTYVPFGGGPRMCPGKEYARLEILVFMHNVVTKFKWEKILPDE-NIVVNPMPV    466
CYP716A79       DGSGPAPYTFVPFGGGPRMCPGKEYARLEILVFMYHIVKRFKWEKVLPNE-KVIVNPMPI    465
CYP716A12       EGNGPAPYTFVPFGGGPRMCPGKEYARLEILVFMHNLVKRFKWEKVIPDE-KIIVDPFPI    465
CYP716A179      EGNGPAPYTFVPFGGGPRMCPGKEYARLEILVFMHNLVKRFKWEMLIPEE-KIVVDPLPM    467
CYP716A147      EGNGPAPYTFVPFGGGPRMCPGKEYARLEILVFMHNVMKRFNWEKLLPDE-KIIVDPLPM    464
CYP716A75       DGSGPAPYTFVPFGGGPRMCPGKEYARLEILVFMHNLVKRFRWEKLIPDE-KIVVNPMPV    473
CYP716A140      EGAGPAPFTFVPFGGGPRMCPGKEYARVEILVFMHHLVKRFKWEKIIPNE-KIVVNPMPC    465
CYP716A15       EGNGPAPYTFVPFGGGPRMCPGKEYARLEILVFMHNVVKRFKWDKLLPDE-KIIVDPMPM    466
CYP716A83       EGTGPPPFTFVPFGGGPRMCPGKEYARLEILVFIHNLVKRYKWEKIIPDE-KIIVNPMPI    467
CYP716A52v2     DGAGPPPFSFVPFGGGPRMCPGKEYARLEILVFMHHLVKRFKWEKVIPDE-KIVVNPMPI    467
CYP716A244      EGAGPPPYSFVPFGGGPRMCPGKEYARLEILVFMYHVVKRFKWEKVIPDE-KIVVNPMPI    467
CYP716A44       EGSGPAPYTFVPFGGGPRMCPGKEYARLEILVFMHHLVKRFKFEKIIPHE-KIIVNPMPI    460
CYP716AL1       DGSGPAPYTFVPFGGGPRMCPGKEYARLEILVFMHNLVKRFKWEKIIPNE-KIVVDPMPI    465
CYP716A48       EGSGPAPYTFVPFGGGPRMCPGKEYARLEILVFMHHLVKRFKWEKLIPDE-KIVVDPMPI    465
                :* ** *::::*********:*:**:***::::::::.::.:: : *.* :::*:*:*

CYP716A80       PGNGLPIRIFPHFGS-----    477
CYP716A1        PDKGLPIRIFPQSGS-----    479
CYP716A110      PAKGLPIRLEPQVCGS----    482
CYP716A79       PEHGLPVRLFPHPQTAVAGS    485
CYP716A12       PAKDLPIRLYPHKAGS----    481
CYP716A179      PANDLPIRLYPHDTGS----    483
CYP716A147      PAKGLPVRLLPHKPGS----    480
CYP716A75       PEKGLPIRLFSYDAGS----    489
CYP716A140      PAKGLPIRLYPHKAGS----    481
CYP716A15       PAKGLPVRLHPHKPGS----    482
CYP716A83       PAKGLPIRLFPHKAGS----    483
CYP716A52v2     PANGLPVRLFPHKAGS----    483
CYP716A244      PASGLPVRLFPHKAGS----    483
CYP716A44       PANGLPLRLYPHHHNPGS--    478
CYP716AL1       PEKGLPVRLYPHINAGS---    482
CYP716A48       PAKGLPIRLYPLNAGS----    481
                *  .**:*:
```

Figure 5.4. Multiple sequence alignment of CYP716As. See next page for full legend.

The plasmids were transformed into *S. cerevisiae* BY4741, and the resulting strains (MD-OAX series) were analysed for triterpenoid production by GC-MS. The OD_{600} at the end of the production run (final OD_{600}) was measured, and growth curves were generated using the same microplate assay as before (see Chapter 4). A strain expressing *AaBAS* alone (MD-BA14) and the MD-N1 control strain were also analysed. The experiment was performed twice and both repeats were in good agreement. In the first experiment, the GC-MS analysis was performed with a 10:1 split ratio for all samples but titres could not be calculated for some samples because the peak areas fell outside of the linear region of the standard curve. The second experiment was therefore performed with additional split ratios (5:1 or 25:1) to allow titres to be calculated. Therefore, only the data from the second experiment is presented here. Nevertheless, some peaks were so small that titres still could not be calculated.

Titres were calculated for ergosterol, β-amyrin, erythrodiol and oleanolic acid, but could not be calculated for oleanolic aldehyde due to the lack of a standard. Instead, normalised peak areas were used to compare oleanolic aldehyde abundance between different strains. Notably, these cannot be used to compare the levels of oleanolic aldehyde with the other triterpenoids. Different analytes can give different signal intensities when analysed by GC-MS, so unless two analytes are known to give very similar signals, their relative abundances cannot be determined using peak areas alone. This was observed when analysing erythrodiol and oleanolic acid standards, where injecting the same amount of each gave noticeably different peak areas (see section 3.4, Figure 3.5 A). Calculating titres from standard curves accounts for such differences in signal intensity, allowing the abundances of different analytes to be

←

Figure 5.4. Multiple sequence alignment of CYP716As. Protein sequences were aligned using Clustal Omega (www.ebi.ac.uk/Tools/msa/clustalo/) with default parameters. Conserved sequences are highlighted. Green: transmembrane domain predicted by the TMHMM server (http://www.cbs.dtu.dk/services/TMHMM); purple: proline cluster that forms a hinge between the transmembrane and cytoplasmic domains (P450 consensus PPxP); light blue: A/G-G-X-D/E-T-T/S motif (proton transfer groove); blue: ExxR motif; red: PERF motif; orange: FxxGxRxCxG motif (haem binding loop). See Table 5.1 for accession numbers. The sequences contained an additional 'GS' at the C-terminus but were otherwise identical to the GenBank sequences.

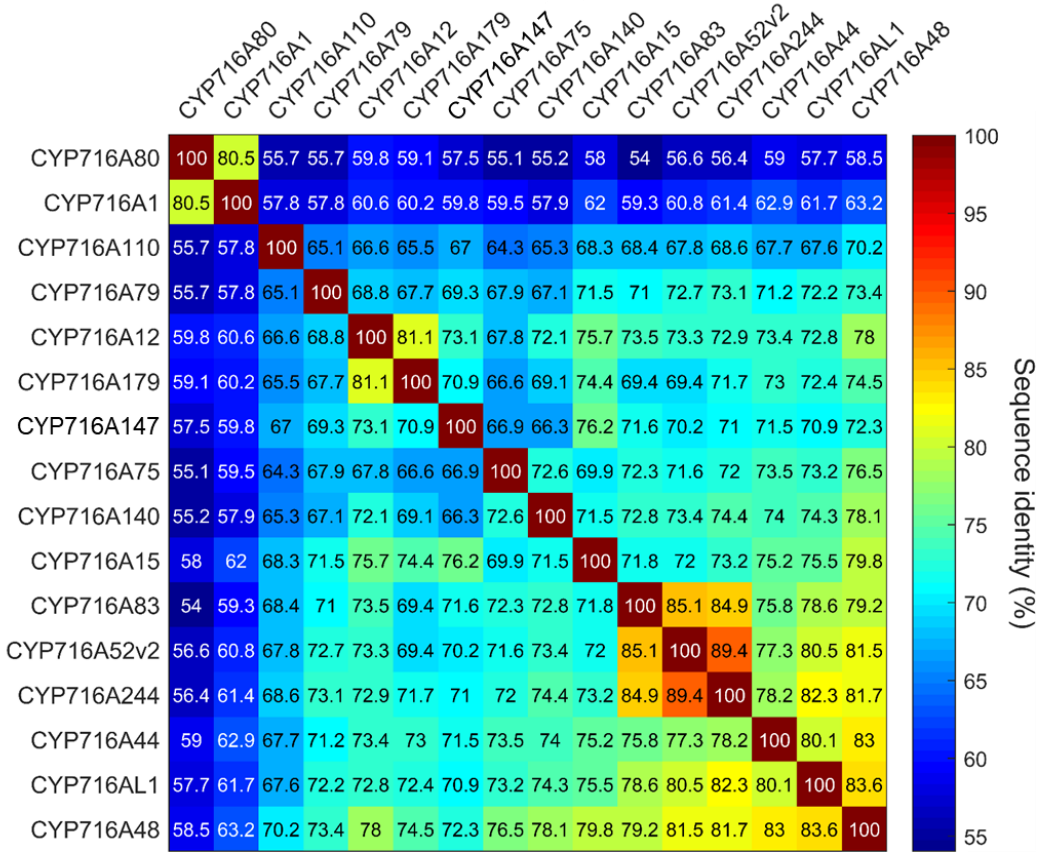


Figure 5.5. Percent identity matrix of BAS sequence identities. Protein sequences were aligned using ClustalOmega (www.ebi.ac.uk/Tools/msa/clustalo/) with default parameters to obtain a percent identity matrix (PIM). All proteins except CYP716A80 and CYP716A1 had sequence identities > 65 %. For accession numbers see Table 5.1.

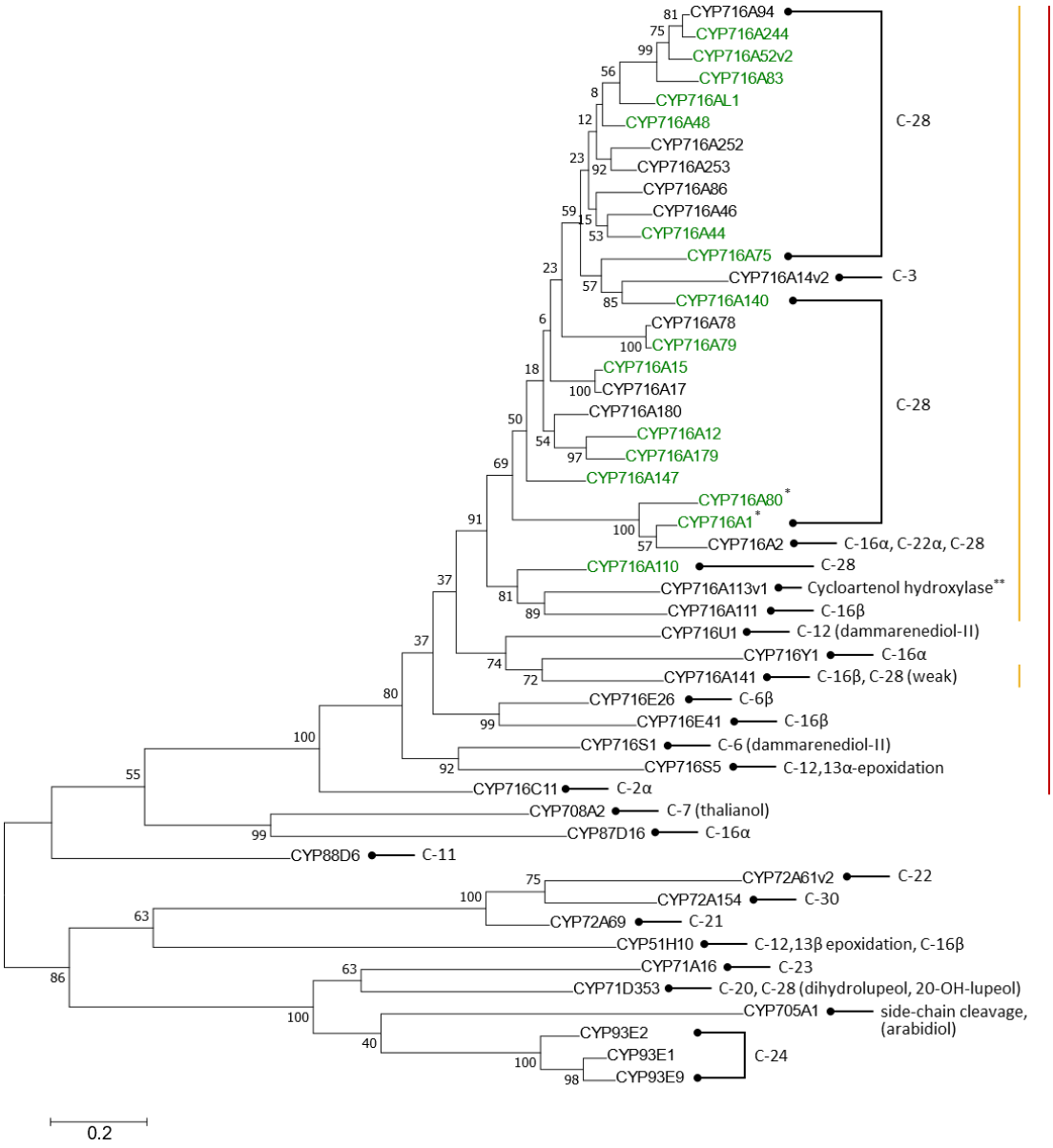

CYP716A94
CYP716A244
CYP716A52v2
CYP716A83
CYP716AL1
CYP716A48
CYP716A252
CYP716A253
CYP716A86
CYP716A46
CYP716A44
CYP716A75
CYP716A14v2
CYP716A140
CYP716A78
CYP716A79
CYP716A15
CYP716A17
CYP716A180
CYP716A12
CYP716A179
CYP716A147
CYP716A80*
CYP716A1*
CYP716A2
CYP716A110
CYP716A113v1
CYP716A111
CYP716U1
CYP716Y1
CYP716A141
CYP716E26
CYP716E41
CYP716S1
CYP716S5
CYP716C11
CYP708A2
CYP87D16
CYP88D6
CYP72A61v2
CYP72A154
CYP72A69
CYP51H10
CYP71A16
CYP71D353
CYP705A1
CYP93E2
CYP93E1
CYP93E9
C-28
C-3
C-28
C-16α, C-22α, C-28
C-28
Cycloartenol hydroxylase**
C-16β
C-12 (dammarenediol-II)
C-16α
C-16β, C-28 (weak)
C-6β
C-16β
C-6 (dammarenediol-II)
C-12,13α-epoxidation
C-2α
C-7 (thalianol)
C-16α
C-11
C-22
C-30
C-21
C-12,13β epoxidation, C-16β
C-23
C-20, C-28 (dihydrolupeol, 20-OH-lupeol)
side-chain cleavage, (arabidiol)
C-24
0.2

Figure 5.6. Phylogenetic analysis of CYP716As and other P450s involved in triterpenoid and sterol biosynthesis. The phylogenetic tree was generated from protein sequences by using the maximum likelihood method and bootstrapping for 1000 replicates. Bootstrap values (%) are displayed at each node. The CYP716As tested in the present study are highlighted green (see Table 5.1 for accession numbers), and the function of each enzyme is annotated. The scale bar represents the number of substitutions per site. Red bar: CYP716 family; yellow bar: CYP716A subfamily; * multifunctional CYP716A; ** putative activity. CYP716A2: *Arabidopsis thaliana* a/ß-amyrin oxidase (BAU61505); CYP716A14v2: *Artemisia annua C*-3 oxidase (AHF22083); CYP716A17: *Vitis vinifera C*-28 oxidase (BAJ84107); CYP716A46: *Solanum lycopersicum C*-28 oxidase (XP_004243906.1); CYP716A78: *Chenopodium quinoa C*-28 oxidase (ANY30853.1); CYP716A86: *Centella asiatica C*-28 oxidase (AOG74831.1); CYP716A111: *Aquilegia coerulea C*-16β hydroxylase (APG38190.1); CYP716A141: *Platycodon grandiflorus C*-16β hydroxylase (AOG74838.1); CYP716A113v1: *Aquilegia coerulea* putative cycloartenol hydroxylase (AOG74849.1); CYP716A252: *Ocimum basilicum C*-28 oxidase (AFZ40057); CYP716A253: *Ocimum basilicum C*-28 oxidase (AFZ40058); CYP716A180: *Betula platyphylla C*-28 oxidase (AHL46848); CYP716A94: *Kalopanax septemlobus C*-28 oxidase (ALO23117.1); CYP716C11: *C. asiatica C*-2α hydroxylase (AOG74835.1); CYP716E26: *S. lycopersicum C*-6β hydroxylase (XP_004241821.1); CYP716E41: *S. asiatica C*-6β hydroxylase (AOG74834.1); CYP716S5: *Platycodon grandiflorus C*-12,13α-epoxidase (AOG74839.1); CYP716Y1: *Bupleurum falcatum C*-16α hydroxylase (AHF45909.1); CYP93E1: *Glycine max C*-24 hydroxylase (BAE94181.1); CYP93E2: *Medicago truncatula C*-24 hydroxylase (ABC59085.1); CYP93E9: *Phaseolus vulgaris C*-24 hydroxylase (AIN25421); CYP51H10: *Avena strigosa* multifunctional oxidase (ABG88961.1); CYP72A61v2: *M. truncatula C*-22 hydroxylase (BAL45199); CYP72A69: *G. max C*-21 hydroxylase (BAW35010.1); CYP72A154: *Glycyrrhiza uralensis C*-30 oxidase (BAL45207); CYP87D16: *Maesa lanceolata C*-16α hydroxylase (AHF22090.1); CYP88D6: *G. uralensis C*-11 oxidase (BAG68929); CYP716U1: *Panax ginseng C*-12 hydroxylase (DM-II) (AEY75212); CYP716S1: *P. ginseng C*-6 hydroxylase (PPD) (AFO63031); CYP705A1: *Arabidopsis thaliana*, arabidiol side-chain cleavage (NP_193268); CYP708A2: *A. thaliana C*-7 hydroxylase (Q8L7D5); CYP71A16: *A. thaliana C*-23 hydroxylase (NP_199073); CYP71D353: *Lotus japonicus C*-20 hydroxylase, *C*-28 oxidation (AHB62239.1).

compared (see Chapter 3 for more information). The “total triterpenoid titre” was also calculated, and is the sum of the titres of ergosterol, β-amyrin, erythrodiol and oleanolic acid; oleanolic aldehyde was not included in this metric because its titres could not be calculated.

Ergosterol and β-amyrin titres were similar for all strains, ranging from 5.6 to 7.5 mg/L for ergosterol (Figure D) and 7.1 to 10.4 mg/L for β-amyrin (Figure 5.7 C). The exception was MD-N1, which accumulated 15 mg/L ergosterol and, as expected, produced no β-amyrin. Similarly, the final OD_{600} was similar across all strains except MD-N1, which reached a considerably higher OD_{600} (Figure 5.7 E). The increased OD_{600} and ergosterol titre of MD-N1 compared with the CYP716A strains is similar to what was observed in Chapter 4, where MD-N1 grew faster and accumulated more ergosterol than the strains expressing a BAS. Notably, β-amyrin titres, ergosterol titres and growth rates were very similar between the CYP716A strains and the strain expressing AaBAS alone (Figure 5.7 C-E).

In contrast with the β-amyrin and ergosterol titres, there were large differences in levels of erythrodiol, oleanolic aldehyde and oleanolic acid between the strains (Figure 5.7 A-B). Oleanolic acid titres ranged from 14.1 mg/L to 2.08 mg/L (a 6.8-fold difference), while erythrodiol titres ranged from 20.9 mg/L to 1.12 mg/L (an 18.7-fold difference). However, for a number of strains the oleanolic acid and erythrodiol peak areas fell below the linear regions of the respective standard curves, preventing titres from being calculated. As such, the actual fold difference in titres is expected to be much greater. Oleanolic aldehyde levels also varied and for some strains, erythrodiol, oleanolic aldehyde and/or oleanolic acid were not observed. Together, this reveals the variation that can occur across homologues that catalyse the same reaction, and highlights the importance of the choice of homologue.

CYP716AL1 accumulated the most oleanolic acid at 14.1 mg/L, and also produced 6.7 mg/L erythrodiol (2.1-fold lower than oleanolic acid) (Figure 5.7 A), representing 41 % and 19 % of total triterpenoids, respectively (Table 5.2). CYP716A12 produced 10.9 mg/L oleanolic acid and 4.6 mg/L erythrodiol (2.4-fold lower), corresponding to 36 % and 15 % of the total triterpenoids. Both accumulated significant amounts of oleanolic aldehyde compared with the other homologues, with CYP716A12 producing slightly less than CYP716AL1 (Figure 5.7 B). As such, both had similar product profiles, favouring the production of oleanolic acid over erythrodiol, with CYP716AL1

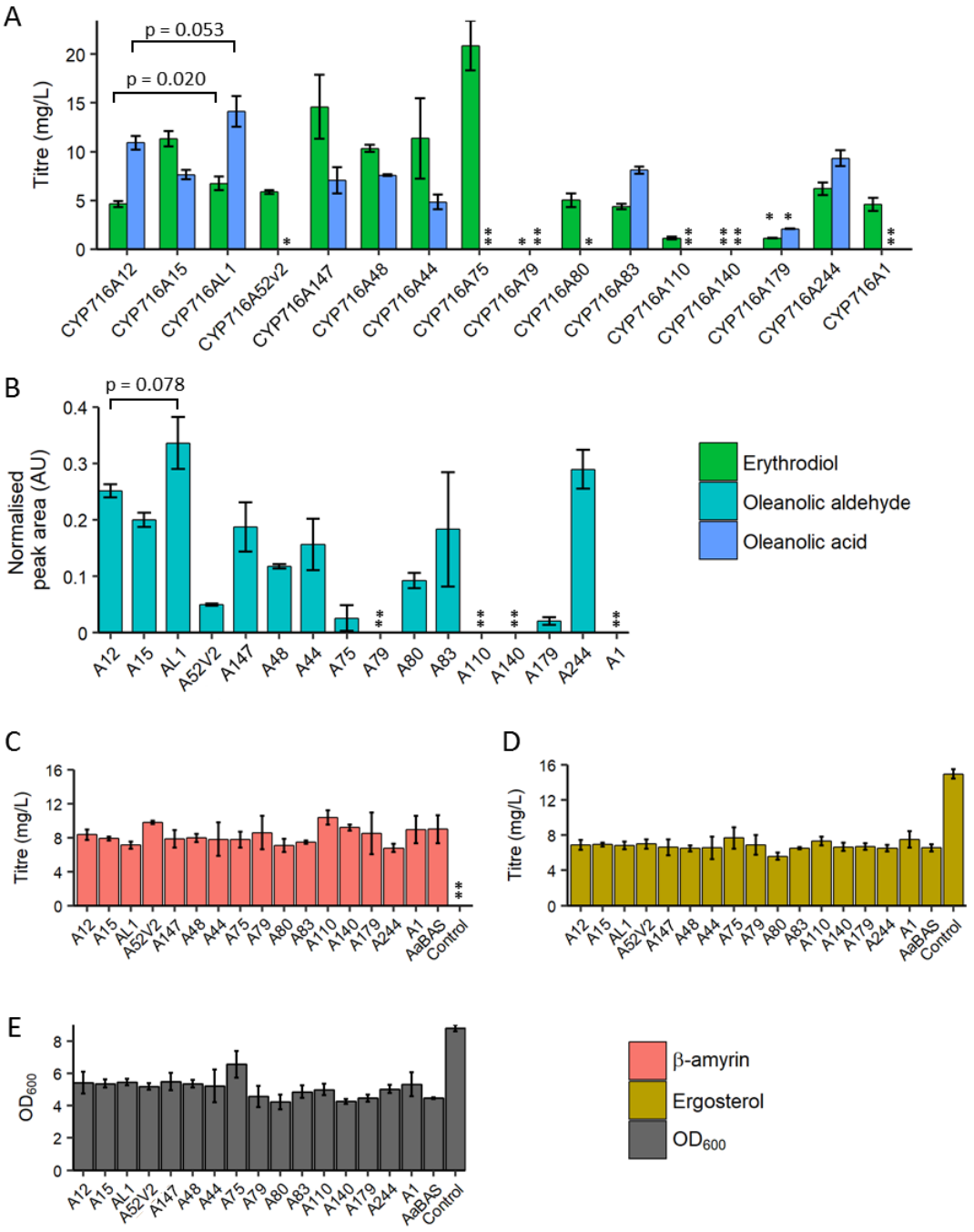


Figure 5.7. Comparison of CYP716As. (A) Erythrodiol and oleanolic acid titres. **(B)** Oleanolic aldehyde normalised peak areas. **(C)** β-amyrin titres. **(D)** Ergosterol titres. **(E)** OD_{600}. Metabolites were measured by GC-MS, and OD_{600} was measured at the end of the production run (96 h). Normalised peak areas are shown for oleanolic aldehyde because a standard was not available, preventing titres from being calculated. While all strains except MD-N1 produced similar amounts of β-amyrin and ergosterol and reached a similar OD_{600}, erythrodiol, oleanolic aldehyde and oleanolic acid levels varied widely. Control: MD-N1 (empty vector). *, titre could not be calculated as the peak area fell below the linear region of the standard curve; **, no peak observed. Shown is the mean ± 1 standard deviation for 3 biological replicates. P-values calculated using a Welch (unequal variances) two-sample t-test in R.

having an overall higher activity. Thus, the comparative analysis immediately identified a CYP716A with greater activity than the homologue commonly used in metabolic engineering projects.

Similar to CYP716AL1 and CYP716A12, CYP716A83 and CYP716A244 also accumulated oleanolic acid in greater abundance than erythrodiol (Figure 5.7 A). CYP716A83 produced 8.1 mg/L oleanolic acid and 4.4 mg/L erythrodiol (a 1.8-fold difference), and accumulated a moderate amount of oleanolic aldehyde compared with the other homologues. Meanwhile, CYP716A244 produced 9.3 mg/L oleanolic acid and 6.2 mg/L erythrodiol (a 1.5-fold difference), and accumulated higher amounts of oleanolic aldehyde compared to the other CYP716As. Oleanolic acid represented 31 % and 32 % of total triterpenoids for CYP716A83 and CYP716A244.

The newly identified P450s (CYP716A147 and CYP716A48) produced similar amounts of oleanolic acid at 7.1 mg/L and 7.6 mg/L, respectively (Figure 5.7 A). However, CYP716A147 produced two-fold more erythrodiol than oleanolic acid at

Table 5.2. Percentage accumulation of triterpenoids

Enzyme	Ergosterol	β-amyrin	Erythrodiol	Oleanolic acid
CYP716A12	22.3 ± 0.9	27.2 ± 0.7	15 ± 0.3	35.5 ± 0.1
CYP716A15	20.6 ± 0.4	23.3 ± 1.2	33.5 ± 1	22.6 ± 0.8
CYP716AL1	19.6 ± 0.6	20.5 ± 0.6	19.4 ± 0.4	40.5 ± 0.9
CYP716A52v2	30.9 ± 1.5	43.2 ± 1	25.9 ± 0.5	0
CYP716A147	18.4 ± 1	21.9 ± 1.1	40.2 ± 1.7	19.5 ± 0.4
CYP716A48	20 ± 0.3	24.6 ± 0.6	31.9 ± 0.3	23.4 ± 0.6
CYP716A44	21.7 ± 1.5	25.6 ± 0.4	36.6 ± 3.4	16.1 ± 1.6
CYP716A75	21.1 ± 0.6	21.4 ± 0.3	57.5 ± 0.4	0
CYP716A79	44.8 ± 2.4	55.2 ± 2.4	0	0
CYP716A80	31.7 ± 1.4	40.1 ± 0.3	28.2 ± 1.2	0
CYP716A83	24.7 ± 0.3	28.2 ± 0.9	16.5 ± 0.5	30.6 ± 0.8
CYP716A110	39 ± 0.5	55.1 ± 0.6	6 ± 0.6	0
CYP716A140	42 ± 1.3	58 ± 1.3	0	0
CYP716A179	40.6 ± 10.3	48.8 ± 1.3	0	0
CYP716A244	22.7 ± 0.5	23.5 ± 0.2	21.5 ± 0.6	32.3 ± 0.4
CYP716A1	35.8 ± 1.6	42.3 ± 1.5	21.8 ± 1.1	0

14.6 mg/L, the second highest erythrodiol titre overall, while CYP716A48 produced less at 10.3 mg/L. For CYP716A147, erythrodiol and oleanolic acid represented 40 % and 20 % of total triterpenoids, while for CYP716A48 they represented 32 % and 23 % (Table 5.2). In addition, CYP716A147 produced slightly more oleanolic aldehyde (1.6-fold higher) than CYP716A48. Thus, CYP716A147 clearly favours erythrodiol over oleanolic acid production, and CYP716A48 has a more balanced product profile.

In addition to CYP716A147 and CYP716A48, a number of other enzymes also accumulated more erythrodiol than oleanolic acid. Strikingly, CYP716A75 produced 20.9 mg/L erythrodiol, representing 58 % of total triterpenoids, yet accumulated very little oleanolic aldehyde compared with the other strains and no detectable oleanolic acid (Figure 5.7 A-B). CYP716A75 therefore appears to be highly specific for the *C*-28 hydroxylation of β-amyrin. This was the highest triterpenoid titre observed in this study, and more than two-fold greater than the β-amyrin titre of AaBAS alone, indicating that there is a significant amount of metabolic "pull-through" occurring. Similarly, CYP716A52v2 produced 5.9 mg/L erythrodiol (26 % of total triterpenoids) and low quantities of oleanolic aldehyde and oleanolic acid compared with the other homologues. CYP716A15 produced 1.5-fold more erythrodiol than oleanolic acid, with titres of 11.3 and 7.6 mg/L, respectively, while CYP716A44 produced 11.4 mg/L erythrodiol and only 4.8 mg/L oleanolic acid (2.4-fold lower than erythrodiol). Erythrodiol thus represented 34 % and 37 % of total triterpenoids for CYP716A15 and CYP716A44, respectively. Both enzymes produced similar amounts of oleanolic aldehyde.

CYP716A1 only produced erythrodiol (4.6 mg/L); oleanolic aldehyde and oleanolic acid were not observed (Figure 5.7 A-B). This observation contrasts with the results of Yasumoto *et al.*, who observed significant amounts of both erythrodiol and oleanolic acid when expressing CYP716A1 in yeast (Yasumoto et al., 2017). Yasumoto *et al.* observed similar peak areas for erythrodiol and oleanolic acid, while in the present study the erythrodiol peak was of considerable size yet no oleanolic acid peak was visible. Thus, differences in the sensitivities of the analytical methods used are unlikely to explain this discrepancy, as such differences would be unlikely to affect the detection of oleanolic acid but not erythrodiol. However, Yasumoto *et al.* used a CPR from *Lotus japonicus* (*LjCPR1*), whereas the present study used *ATR2* (from *A. thaliana*). As such, differences in electron transfer efficiencies between the CPR and

CYP716A1 might explain the differences in product ratios between these two studies. Interestingly, Yasumoto *et al.* (2016) reported that CYP716A1 was only capable of performing *C*-28 hydroxylation on lupeol, producing betulin, which is equivalent to the hydroxylation of β-amyrin to erythrodiol. They used a diploid yeast strain (INVSc1), expressed three copies of *CYP716A1*, used *LjCPR1* and cultured cells for two days in galactose-supplemented media (Yasumoto et al., 2016). By contrast, the present study used a haploid yeast strain expressing a single copy of *CYP716A1* and the CPR *ATR2*, and cultures were performed for 4 days. These differences could explain the discrepancies between our results.

CYP716A79 and CYP716A110 did not produce any detectable oleanolic aldehyde or oleanolic acid, and accumulated very low quantities of erythrodiol (Figure 5.7 A-B). CYP716A110 produced 1.12 mg/L erythrodiol, while the titre could not be calculated for CYP716A79 because it fell below the linear region of the standard curve. However, both enzymes were previously reported to oxidise β-amyrin to oleanolic acid (Fiallos-Jurado et al., 2016; Miettinen et al., 2017). These studies used *S. cerevisiae* BY4742, a haploid yeast strain similar to strain BY4741 used in the present study, but overexpressed the *tHMG1* gene, repressed the *ERG7* gene using a methionine-repressible promoter, and added methyl-β-cyclodextrin to the growth medium, which sequestered triterpenoids in the growth medium. These are strategies that have been shown to boost triterpenoid production (Moses et al., 2014b; Fiallos-Jurado et al., 2016; Miettinen et al., 2017). Furthermore, Miettinen *et al.* used a different CPR (*MTR1* from Medicago truncatula instead of *ATR2*), which could have affected P450 activity (Miettinen et al., 2017). These differences could explain why oleanolic acid was detected in those studies yet was not detected here.

CYP716A80 also accumulated significantly more erythrodiol than oleanolic acid, producing 5 mg/L erythrodiol, while the oleanolic acid peak was too small for titres to be calculated (Figure 5.7 A). It also produced a relatively small amount of oleanolic aldehyde compared with the other CYP716As (Figure 5.7 B). However, two additional products (U3 and U4) were also observed (Figure 5.8 A). These had retention times of 18.48 min (U3) and 18.88 min (U4), and had slightly larger peaks than oleanolic acid. Khakimov *et al.* also reported additional products with similar retention times for this enzyme, and reasoned that both were β-amyrin derivatives, the first containing two hydroxyl groups and the other being similar but containing an additional double

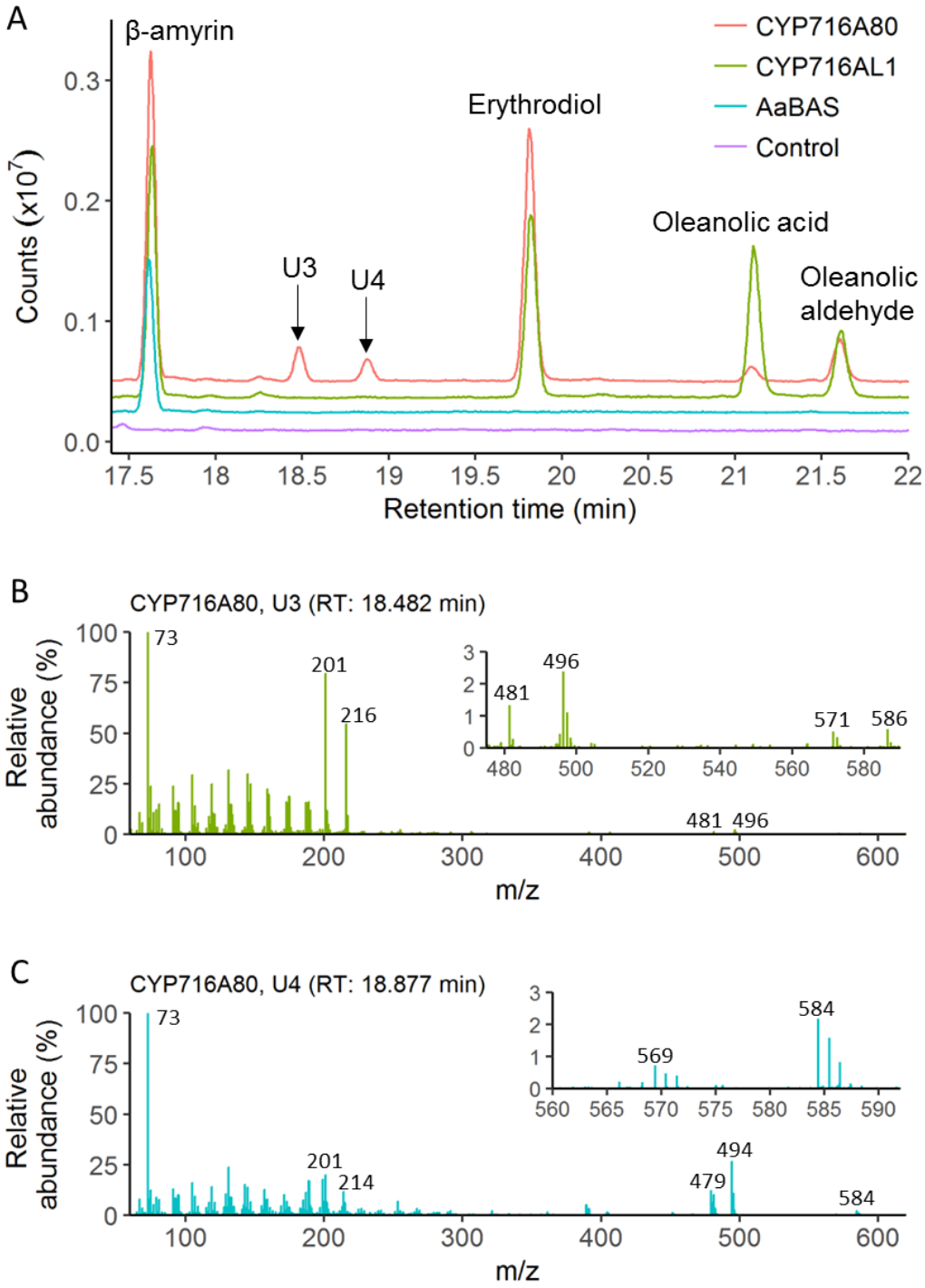


Figure 5.8. CYP716A80 produces additional products. (A) Total ion chromatogram overlay of the strains expressing the oleanolic acid pathways containing CYP716A80 and CYP716AL1, AaBAS alone, or the empty vector control (MD-N1). The two additional peaks observed for CYP716A80 (U3 and U4) are labelled. **(B-C)** Mass spectra of peaks U3 and U4, with background mass spectra (corresponding to 19.049 - 19.545 min of each sample) subtracted. Both are consistent with them being β-amyrin derivatives.

bond (Khakimov et al., 2015). Indeed, the mass spectrum of U3 contains a fragment ion with an m/z of 216 (Figure 5.8 B), which is characteristic of β-amyrin derivatives that contain a hydroxyl group in their diene fragment (see Chapter 3). Furthermore, a low abundance peak with an m/z of 496 could result from the loss of a TMSOH (90 Da) from the parent ion, and a low abundance peak with an m/z of 586 is indeed visible in the mass spectrum (Figure 5.8 B). In this case, the parent ion of U3 would have the same m/z as the erythrodiol parent ion (586). This would indicate that U3 and erythrodiol are structural isomers, with U3 being derived from β-amyrin by a single hydroxylation, supporting the hypothesis that U3 contains two hydroxyl groups.

The mass spectrum of U4 contained peaks with m/z values of 584, 569, 494 and 479, which could represent the parent ion, the loss of a TMSOH (90 Da) from the parent, and/or the loss of a methyl group (15 Da) (Figure 5.8 C). This indicates that U4 also contains a hydroxyl group. Furthermore, the mass spectrum contains ions with m/z values of two lower than the corresponding ions for U3 (e.g. 584 vs. 586, 569 vs. 571, 494 vs. 496 and 479 vs. 481). This is consistent with it containing an additional double bond as proposed by Khakimov *et al.* (Khakimov et al., 2015), because the addition of a double bond is accompanied by the loss of two hydrogen atoms, decreasing the m/z by two. Comparison with the mass spectrum of erythrodiol (see Figure 3.4 B and Table 3.1 in Chapter 3) shows that the erythrodiol parent ion (m/z 586) has an m/z of two higher than the U4 parent ion, and erythrodiol has fragment ions with m/z values of 496 and 481, while U4 has fragments ions with m/z values of 494 and 479 (two m/z values lower). This is consistent with U4 having two hydroxyl groups (like erythrodiol) but an additional double bond. U4 does not contain the fragment ion with an m/z of 216, which is characteristic of β-amyrin derivatives containing a hydroxyl in their diene fragment. However, this would make sense if the additional double bond was present in the diene fragment. In this case, a fragment ion with m/z 214 would be expected, and a peak with this m/z value is indeed observed (Figure 5.8 C).

No oxidised products were observed for CYP716A140 (Figure 5.7 A-B). However, it was previously reported to oxidise β-amyrin, as well as 12,13α-epoxy-β-amyrin and 16β-hydroxy-β-amyrin, with the latter being the favoured substrate (Miettinen et al., 2017). It is likely that this enzyme simply showed low activity in this study, and its products could not be detected using the current method. Finally, while erythrodiol, oleanolic aldehyde and oleanolic acid were all observed for CYP716A179, these were

in low abundance (Figure 5.7 A-B), with erythrodiol and oleanolic acid titres of only 1.13 mg/L and 2.08 mg/L, respectively. These represent the joint lowest erythrodiol titres (with CYP716A110, which produced 1.12 mg/L) and the lowest oleanolic acid titre of the CYP716As.

In addition to the expected products, two further peaks were observed for all CYP716As, the first eluting ~ 0.6 min after β-amyrin and the second ~ 0.4 min after erythrodiol (Figure 5.9 A). Most of the BASs tested in the BAS comparative analysis, including AaBAS, had an additional small peak that also eluted ~ 0.6 min after β-amyrin, and appeared to be lupeol based on their mass spectra (see Figure 4.14 of Chapter 4). For BvBAS, this peak was much larger and was suggested to be lupeol based on comparison of its mass spectrum with those reported in the literature. For the current experiment, the strain expressing *AaBAS* alone had a small putative peak at this position (Figure 5.9 A), although it was smaller than the corresponding peak observed for AaBAS in Chapter 4 (Figure 4.14). The corresponding peak in the CYP716A strains was larger, perhaps suggestive of metabolic pull-through. However, in both cases the signal was weak enough to make interpretation of the mass spectra difficult (Figure 5.9 B-C). Therefore, the first additional peak potentially corresponds to lupeol, but further analysis is required to confirm this. If the first peak is indeed lupeol, then the second additional peak may be betulin, the *C*-28 hydroxy derivative of lupeol. Betulin is the lupane equivalent of erythrodiol, which is the *C*-28 hydroxy derivative of β-amyrin, and this is therefore consistent with it having a similar retention time to erythrodiol. However, as with the putative lupeol peak, it was in low abundance and the mass spectrum was difficult to interpret. As such, the identities of these peaks cannot be confirmed.

Interestingly, there was no noticeable substrate depletion in the CYP716A strains. All accumulated similar amounts of β-amyrin to the strain expressing *AaBAS* alone, despite many producing significant amounts of erythrodiol and oleanolic acid (Figure 5.7 C). Indeed, CYP716A75 produced 7.8 mg/L β-amyrin and 20.9 mg/L erythrodiol, for a total triterpenoid content of 28.7 mg/L. By contrast, the AaBAS strain produced 9 mg/L β-amyrin, a > 3-fold lower total triterpenoid titre than CYP716A75. Thus, expression of the CYP716A enzymes did not deplete the cellular β-amyrin pool. This may be indicative of a metabolic “pull-through” effect, where expression of downstream enzymes increases the flux through the upstream parts of the pathway.

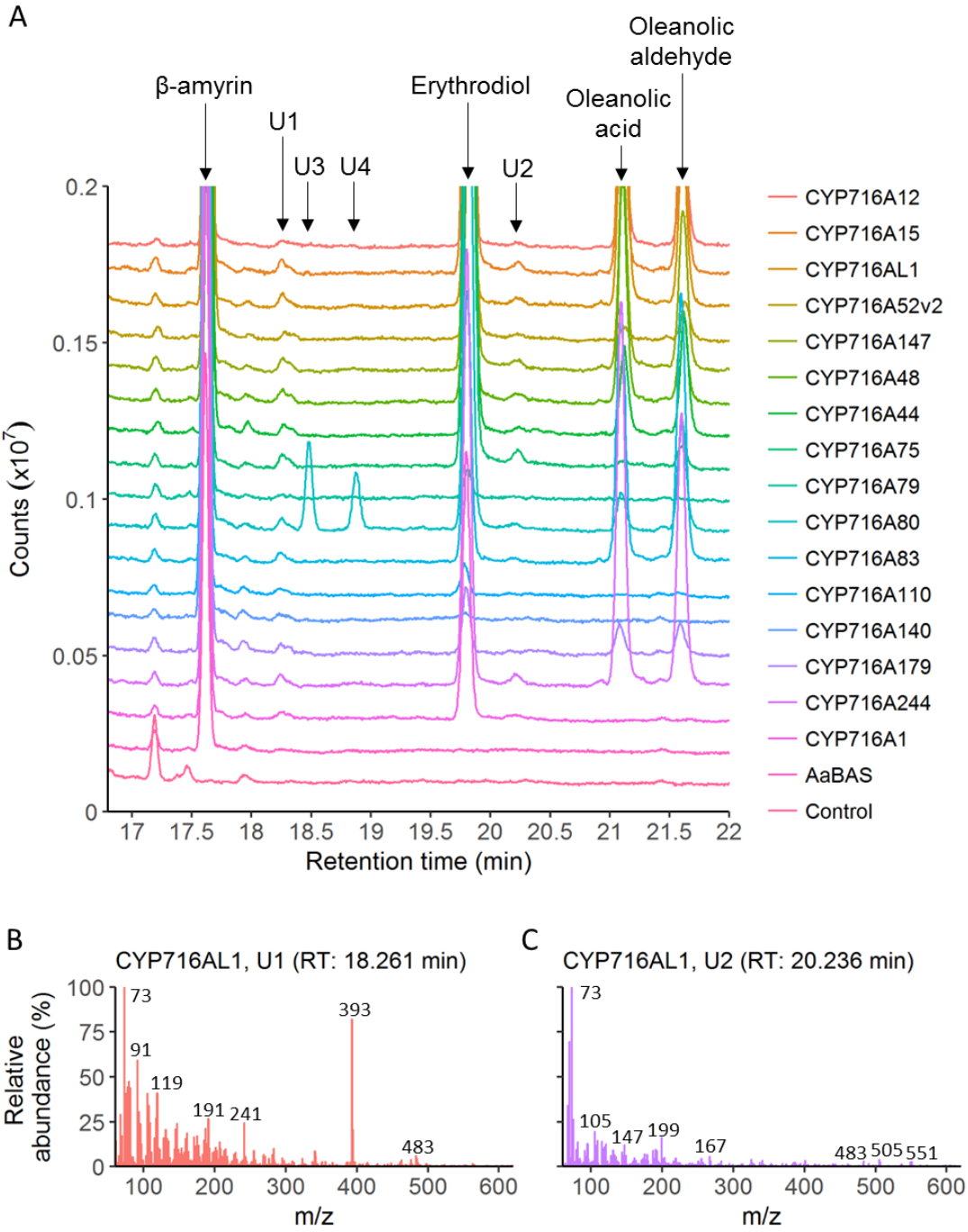


Figure 5.9 Additional triterpenoid produced by the other CYP716As. (A) Overlay of total ion chromatograms showing the two additional peaks (U1 and U2) observed for many CYP716As. Peaks U3 and U4, which were only observed for CYP716A80, are also labelled. **(B-C)** Mass spectra of peaks U1 and U2 from CYP716AL1, with the background spectra (corresponding to 18.601 - 19.000 min of each sample) subtracted. Peak U1 was also observed for AaBAS, while U2 was unique to the strains expressing a CYP716A.

In strains expressing only a BAS, β-amyrin titres may be constrained by toxicity or inhibition of the BAS when cellular β-amyrin levels reach a certain threshold, preventing it from accumulating further. Utilisation of β-amyrin by the CYP716As might relieve this inhibition and allow more β-amyrin to be produced by the BAS, the result being that β-amyrin is not depleted by CYP716A expression but instead remains at a constant level.

5.4 Growth of CYP716A strains

The growth curves showed that all CYP716A strains grew more slowly in galactose-supplemented media than the empty vector control (MD-N1) (Figure 5.10 A). By contrast, all grew similarly in glucose-supplemented media (Figure 5.10 B), which was also seen in the BAS comparative analysis (see Chapter 4). The relative final cell densities from the growth curves generally agreed with the relative final OD_{600} values from the production runs. The strains can be divided into two groups based on the shape of their growth curves in galactose-supplemented media. The first group grew similarly to the strain expressing AaBAS alone, having a relatively short lag phase similar to MD-N1 (~ 10 h) and slow exponential growth (Figure 5.10 A). It included CYP716A52v2, CYP716A79, CYP716A110, CYP716A140, CYP716A179 and CYP716A1. The second group, containing the remaining strains, had a longer lag phase of ~ 30 h but a fast exponential phase.

All CYP716As in the first group produced little or no oleanolic aldehyde and oleanolic acid, and most produced little or no erythrodiol (Figure 5.7 A-B). In fact, the first group contained all the CYP716As that produced little of any oxidised product (CYP716A79, CYP716A110, CYP716A140, and CYP716A179). The exceptions were CYP716A52v2 and CYP716A1, which produced significant amounts of erythrodiol (5.8 mg/L and 4.6 mg/L, respectively). However, CYP716A52v2 accumulated very little oleanolic aldehyde and oleanolic acid, and CYP716A1 produced only erythrodiol. Thus, these P450s produced low levels of oxidised triterpenoids.

Most enzymes in the second group produced significantly more oleanolic aldehyde and oleanolic acid than those in the first group, and generally had higher activities, producing more oxidised triterpenoids (Figure 5.7 A-B). The exceptions to this were CYP716A75 and CYP716A80. The latter accumulated similar amounts of erythrodiol, oleanolic aldehyde and oleanolic acid to CYP716A52v2. However, it also produced two additional products in significant amounts (Figure 5.8 A), and the amount of total

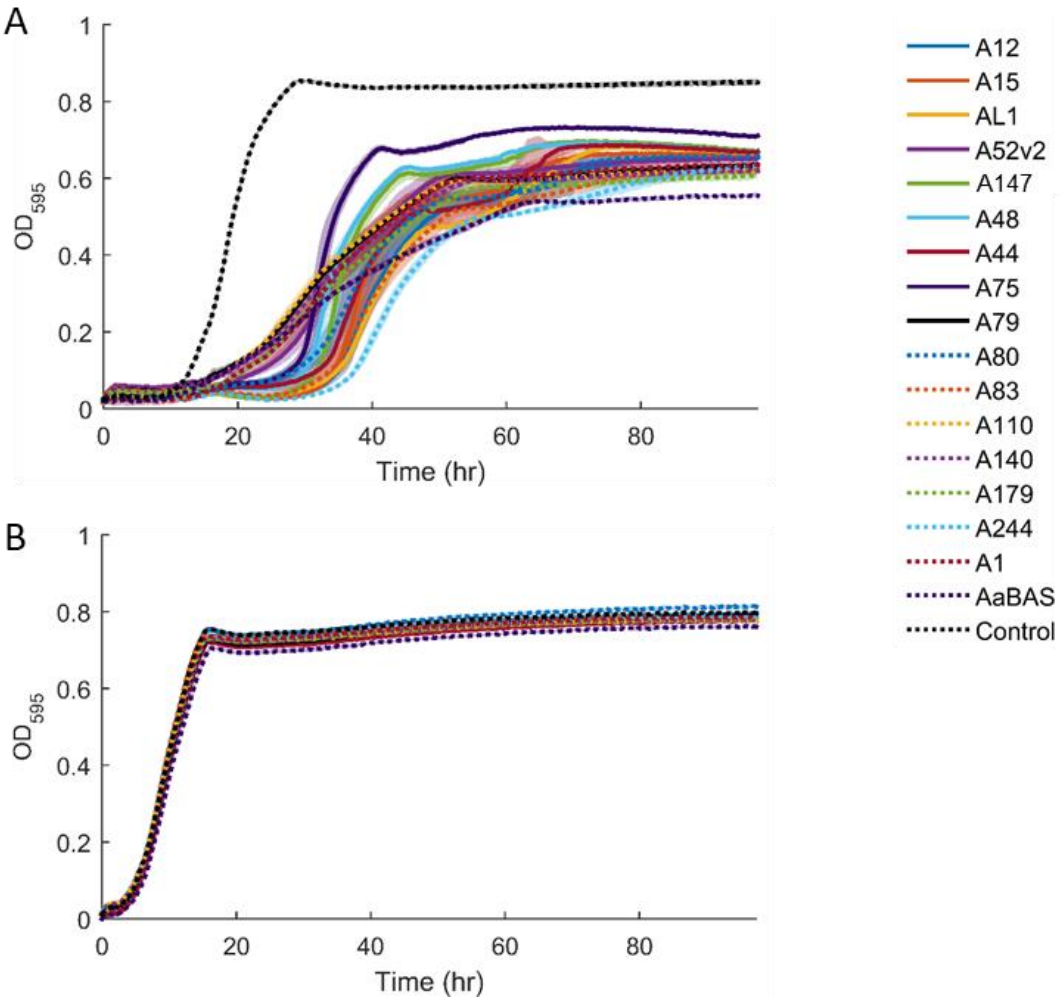


Figure 5.10. Growth of CYP716A strains in a 96-well plate assay. Cells were grown in 96-well plates in media supplemented with either galactose **(A)** or glucose **(B)**. OD_{595} was measured every 15 min for 96 h using a Tecan Sunrise plate reader. All triterpenoid-producing strains grew considerably slower than the empty vector control (MD-N1) in galactose, while growth was similar in glucose. Shown is the mean ± 1 standard deviation for 3 biological replicates.

oxidised triterpenoids is therefore probably higher. CYP716A75 produced very low levels of oleanolic aldehyde and no detectable oleanolic acid, similar to the group one P450s (Figure 5.7 A-B). However, it had a high erythrodiol titre and therefore accumulated a large amount of oxidised triterpenoid (Figure 5.7 A, Table 5.2). CYP716A75 grew faster than the other P450s in the second group, entering stationary phase slightly earlier and reaching a higher OD_{595}, and also reached a slightly higher final OD_{600} in the production runs compared with the other strains (Figure 5.8 A).

It was interesting that some CYP716A strains (group one) had similar growth profiles to the strain expressing *AaBAS* alone, while for others (group two) the shape of the growth curve was markedly different (Figure 5.8 A). The group one strains grew slightly faster than AaBAS, having a faster exponential phase, entering stationary phase earlier, and reaching a higher final OD_{595}. This suggests that there was little additional burden to expressing these enzymes, and that there may be a small growth benefit to expressing them. The different shapes of the growth curves between the group 2 strains and the AaBAS strain indicate that these CYP716As also have an impact on cell burden. However, it is unclear from the present data what is causing these differences. Overall, the growth curves, combined with the similar ergosterol titres and final OD_{600} of the CYP716A strains and the AaBAS strain, indicate that there is relatively little additional burden caused by CYP716A expression, and that BAS expression may cause the majority of the cellular burden.

The differences in growth cannot be explained by differences in β-amyrin or ergosterol levels, as these are similar across all strains except MD-N1, as is the final OD_{600} (Figure 5.7 C-E). It is possible that the total amount of oxidised triterpenoids has an impact, as the group one strains all accumulated less oxidised triterpenoids than the group two strains. It is also possible that β-amyrin accumulation in the cells is toxic, and therefore reduces cell growth, while the accumulation of oxidised triterpenoids works to offset this by some unknown mechanism. Triterpenoids are structurally similar to sterols and may therefore insert into cell membranes, which could disrupt their integrity and cause the observed reductions in growth rate. Indeed, analysis of a strain expressing a catalytically inactive AaBAS suggested that the burden observed in the AaBAS strain might be partly due to β-amyrin production, although it remained possible that differences in expression levels between the wild-type and inactive AaBAS variants could have explained the data (see section 4.5).

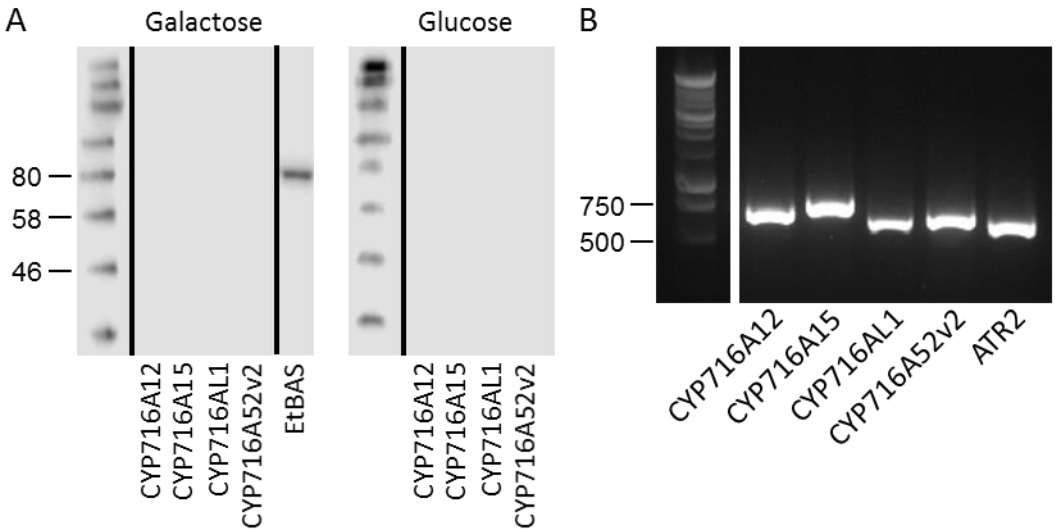


Figure 5.11. Analysis of CYP716A and ATR2 expression by western blot and RT-PCR. Western blots **(A)** and RT-PCR **(B)** were performed on strains expressing 6xHis-tagged CYP716A12, CYP716A15, CYP716AL1, CYP716A52v2 and ATR2. For the western blots, each strain co-expressed a CYP716A and ATR2. A strain expressing His-tagged EtBAS was also analysed as a positive control. No CYP716A or ATR2 protein was detected by western blot, whereas mRNA was detected by RT-PCR, indicating that the genes were expressed. The expected molecular weights of the CYP716As and ATR2 were 54 kDa and 79 kDa, respectively. Molecular weights for the protein and DNA ladders are in kDa and bp, respectively.

Another possible explanation for the observed differences in growth rate could be differences in the expression levels of the enzymes. It is possible that the group one P450s were expressed at low levels, while those in group two were expressed more highly. This could explain why the group one P450s produced lower amounts of oxidised triterpenoids compared with the group two P450s. In addition, it might explain the similar growth profiles of the group one strains to the AaBAS strain, as they could be expressed at such low levels as to not cause any burden, while higher expression of the group two P450s could explain their more substantial effect on growth. However, the group one strains did grow slightly faster than AaBAS, and in fact all strains reached slightly higher ODs than AaBAS (Figure 5.7 E, Figure 5.8 A), which is unlikely to be explained by differences in expression levels alone. Differences in stability of the P450s could also contribute to the differences in growth rate, as misfolding of the P450s (or indeed ATR2) could cause burden on the unfolded protein response, affecting growth.

Western blot analysis of four P450s (CYP716A12, CYP716A15, CYP716AL1 and CYP716A52v2) and ATR2 was unable to detect these enzymes (Figure 5.11 A), while mRNA was detected by RT-PCR (Figure 5.11 B). Furthermore, five of six selected BASs were detected by western blot (see Chapter 4, Figure 4.11 A); AsBAS was not detected. This might be consistent with the P450s and *ATR2* being expressed at lower levels than the BASs. However, further work will be necessary to successfully detect the P450s and ATR2 and confirm the relative protein levels of the BASs, P450s and ATR2.

Although variations in expression level could explain some of the differences in activity between the CYP716As, it would not explain differences in product profiles. These would instead be best explained by differences in the enzymes themselves. Therefore, to properly understand why the CYP716A homologues vary in terms of their product profiles, overall activities and burden, expression levels must be determined and *in silico* analyses performed to assess these differences. While expression levels unfortunately could not be determined, *in silico* analyses were nevertheless performed to investigate the possible causes of the differences in product profiles.

5.5 *In silico* analyses of CYP716A homologues

The CYP716As had a range of activities and product profiles, with CYP716A75 producing almost exclusively erythrodiol and others, such as CYP716AL1, producing significant amounts of oleanolic acid. Furthermore, CYP716A80 accumulated two additional triterpenoids in notable abundance. As such, the multiple sequence alignments generated earlier (Figure 5.4) were analysed to investigate the cause of this variation.

Most of the CYP716As had 64 % to 90 % sequence identities (Figure 5.5) and only displayed β-amyrin *C*-28 oxidase activity, accumulating erythrodiol, oleanolic aldehyde and oleanolic acid (Figure 5.7 A-B). However, CYP716A80 and CYP716A1 had relatively low sequence identities (< 63 %) to the other CYP716As and 81 % identity to each other. Both displayed β-amyrin *C*-28 oxidise activity but CYP716A80 also produced two additional products. Although CYP716A1 only produced erythrodiol, a previous study reported it to have weak *C*-22α hydroxylation activity against α-amyrin, and to oxidise lupeol to an unknown product (Yasumoto et al., 2016). All other CYP716As tested here have only been reported to have *C*-28 oxidase

activity. Thus, these differences in activities appear to correlate well with sequence identities.

The P450 superfamily has very low sequence conservation, and sequence identities can be < 20 %. However, they have a highly conserved tertiary structure, particularly around the catalytic core (Werck-Reichhart and Feyereisen, 2000). The active site contains a haem group, with the iron coordinated by a cysteine residue that is conserved across all P450s (Bak et al., 2011). Important sequence motifs in the core include: the haem binding loop containing the highly conserved FxxGxRxCxG motif, which contains the conserved cysteine; the ERR triad, which is important in stabilising the core structure, and corresponds to the Glu and Arg residues of the EXXR motif and the Arg residue of the PERF motif; and the proton transfer groove (A/GCxD/ETT/S motif) (Werck-Reichhart and Feyereisen, 2000; Bak et al., 2011). Plant P450s are membrane bound, typically localising to the endoplasmic reticulum (ER) (Bak et al., 2011). Such P450s have an N-terminal membrane anchor rich in hydrophobic sequences, often followed by a cluster of basic residues called the halt transfer signal and a cluster of prolines (the PPXP motif) that acts as a hinge between the membrane and cytoplasmic domains of the protein (Werck-Reichhart and Feyereisen, 2000). The N-terminal anchor and the substrate binding regions show the highest sequence variability (Werck-Reichhart and Feyereisen, 2000).

Analysis of the multiple sequence alignment identified the highly conserved haem binding loop (FxxGxRxCxG motif), with all tested CYP716As sharing the sequence FGGGPRMCPG (Figure 5.4). The CYP716As had the consensus sequences E-V-M/L-R and P-S/T/N-R-F/Y, which correspond to the ExxR and PERF motifs that make up the ERR triad. The proton transfer groove had the sequence GGHDTA for all CYP716As, differing from the P450 consensus (A/G-G-X-D/E-T-T/S) only in the final residue (A instead of T/S). The position of the membrane anchors were predicted using the TMHMM server (http://www.cbs.dtu.dk/services/TMHMM), and were located at the N-terminus as expected (Figure 5.4). In addition, a cluster of three proline residues was located just C-terminal to the membrane anchor (sequence PPGxxGxP), which may correspond to the hinge region. The CYP716As typically contained two to three Lys or Arg residues between the membrane anchor and proline cluster, corresponding to the cluster of basic residues that has been reported for P450s (Werck-Reichhart and Feyereisen, 2000). Unfortunately, the low sequence

conservation outside of these regions makes it difficult to identify amino acid variants responsible for the observed differences in product specificity and activity.

5.6 Comparing BASs and CYP716As with different productivities for oleanolic acid production

The CYP716As were co-expressed with *AaBAS*, which was identified as the most productive BAS homologue from the comparative analysis. CYP716AL1 was found to produce the most oleanolic acid in the CYP716A comparative analysis. Thus, the CYP716AL1 strain (MD-OA15) should represent an optimal strain for oleanolic acid production in terms of the BAS and CYP716A homologue expressed. This strain was compared with strains expressing less productive BAS and CYP716A homologues, to determine whether combining the optimal homologues of each enzyme class did indeed boost production. The following strains were analysed: MD-OA15 (expressing *AaBAS* and *CYP716AL1*), MD-OA16 (*AaBAS*, *CYP716A52v2*), MD-OA30 (*SlBAS*, *CYP716AL1*) and MD-OA31 (*SlBAS*, *CYP716A52v2*). SlBAS gave the lowest β-amyrin titres after PgBAS1, while CYP716A52v2 produced low but detectable levels of oleanolic aldehyde and oleanolic acid. Co-expression of these enzymes is therefore expected to result in the production of low but detectable amounts of triterpenoids. By contrast, use of poor performing homologues such as PgBAS1 or CYP716A179 are less likely to result in detectable peaks in the total ion chromatograms, which would prevent comparison of titres with the strains expressing AaBAS and CYP716AL1.

The strains were analysed by GC-MS, and growth curves were generated using a microplate assay as before. For the GC-MS analysis, strains were grown in shake flasks, which may improve growth and possibly production by providing better agitation and aeration. MD-OA31 (*SlBAS*, *CYP716A52v2*) produced very low levels of erythrodiol (1.2 mg/L), oleanolic aldehyde and oleanolic acid, however titres for the latter could not be calculated due to the small peak area (Figure 5.12 A-B). By contrast, MD-OA16, which expressed *AaBAS* and *CYP716A52v2*, produced 5.4-fold more erythrodiol (6.4 mg/L), more oleanolic aldehyde, and small amount of oleanolic acid, which again was not abundant enough for titres to be calculated. It also accumulated substantially more β-amyrin (9.6 mg/L) than MD-OA31 (3.0 mg/L),

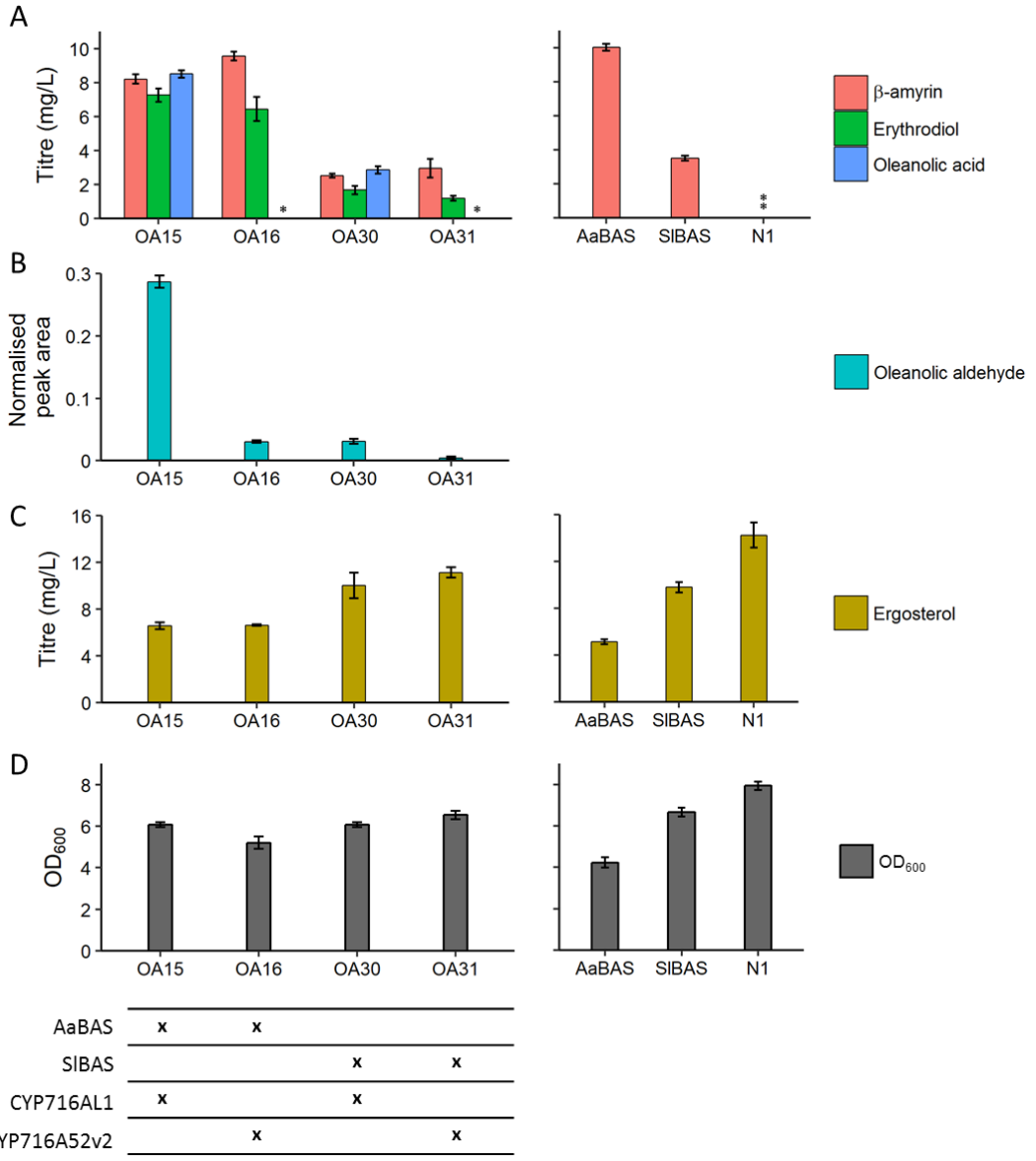


Figure 5.12. Comparison of "optimal" and "suboptimal" oleanolic acid pathways. Strains expressing combinations of BASs and CYP716As with high (*AaBAS*, *CYP716AL1*) or low (*SlBAS*, *CYP716A52v2*) productivities were analysed for triterpenoid production by GC-MS. Control strains expressing the BAS alone or carrying an empty vector (MD-N1) were also analysed. **(A)** β-amyrin, erythrodiol and oleanolic acid titres. **(B)** Oleanolic aldehyde normalised peak areas. **(C)** Ergosterol titres. **(D)** OD_{600} at the end of the production run. The strain expressing the best performing BAS and CYP716A homologues (MD-OA15; *AaBAS* and *CYP716AL1*) produced considerably more triterpenoids than that expressing the poor performing homologues (MD-OA31; *SlBAS* and *CYP716A52v2*). *, titre could not be calculated as the peak area fell below the linear region of the standard curve; **, no peak observed in the TIC. Shown is the mean ± 1 standard deviation for 3 biological replicates.

indicating that the choice of BAS had a significant effect on triterpenoid titres and product ratios by providing a greater β-amyrin substrate pool for the P450. Commensurate with this, the strain expressing *AaBAS* alone produced substantially more β-amyrin than that expressing *SlBAS*, and these strains both produced slightly more β-amyrin than the corresponding oleanolic acid strains expressing the same BAS (Figure 5.12 A).

MD-OA30 (*SlBAS*, *CYP716AL1*) produced a similar amount of erythrodiol (1.7 mg/L), more oleanolic aldehyde and considerably more oleanolic acid (2.9 mg/L) than MD-OA31, while accumulating similar amounts of β-amyrin (2.5 mg/L) (Figure 5.12 A-B). This strain also accumulated significantly less β-amyrin, erythrodiol, oleanolic aldehyde and oleanolic acid than MD-OA15 (*AaBAS*, *CYP716AL1*). Comparison of MD-OA31 with MD-OA15, which represent the “suboptimal” and “optimal” strains in terms of BAS and CYP716A homologue, respectively, showed that MD-OA15 produced 2.8-fold more β-amyrin and 6.1-fold more erythrodiol. The fold difference in oleanolic acid titres could not be calculated because low productivity of MD-OA31 prevented oleanolic acid titres from being determined. However, MD-OA15 accumulated 3-fold more oleanolic acid than MD-OA30. Thus, combining the most productive BAS with the most productive CYP716A greatly improved triterpenoid titres.

The final OD_{600} was similar for all oleanolic acid strains, indicating that changing the BAS and P450 had little effect on the final cell density (Figure 5.12 D). By contrast, the strain expressing *SlBAS* alone reached a higher OD_{600} than *AaBAS*. Ergosterol titres were a little higher in the oleanolic acid strains expressing *SlBAS* compared with *AaBAS*, and were similar to the strains expressing a BAS only (Figure 5.12 C).

The growth in galactose-supplemented media varied depending on the strain (Figure 5.13 A), while all strains grew similarly in glucose-supplemented media (Figure 5.13 B). Notably, the strains expressing *AaBAS* grew faster than those expressing *SlBAS* (Figure 5.13 A). SlBAS grew very slowly, as did the corresponding oleanolic acid strains expressing this BAS (MD-OA30 and 31). MD-OA16, which expressed *AaBAS* and *CYP716A52v2*, grew faster than the *AaBAS*-only strain, while MD-OA15 (*AaBAS*, *CYP716AL1*) had a longer lag phase but a much faster exponential phase than MD-OA16 and the *AaBAS*-only strain (Figure 5.13 A). However, due to relatively large error bounds in the MD-OA16 growth curve, it cannot be determined whether the

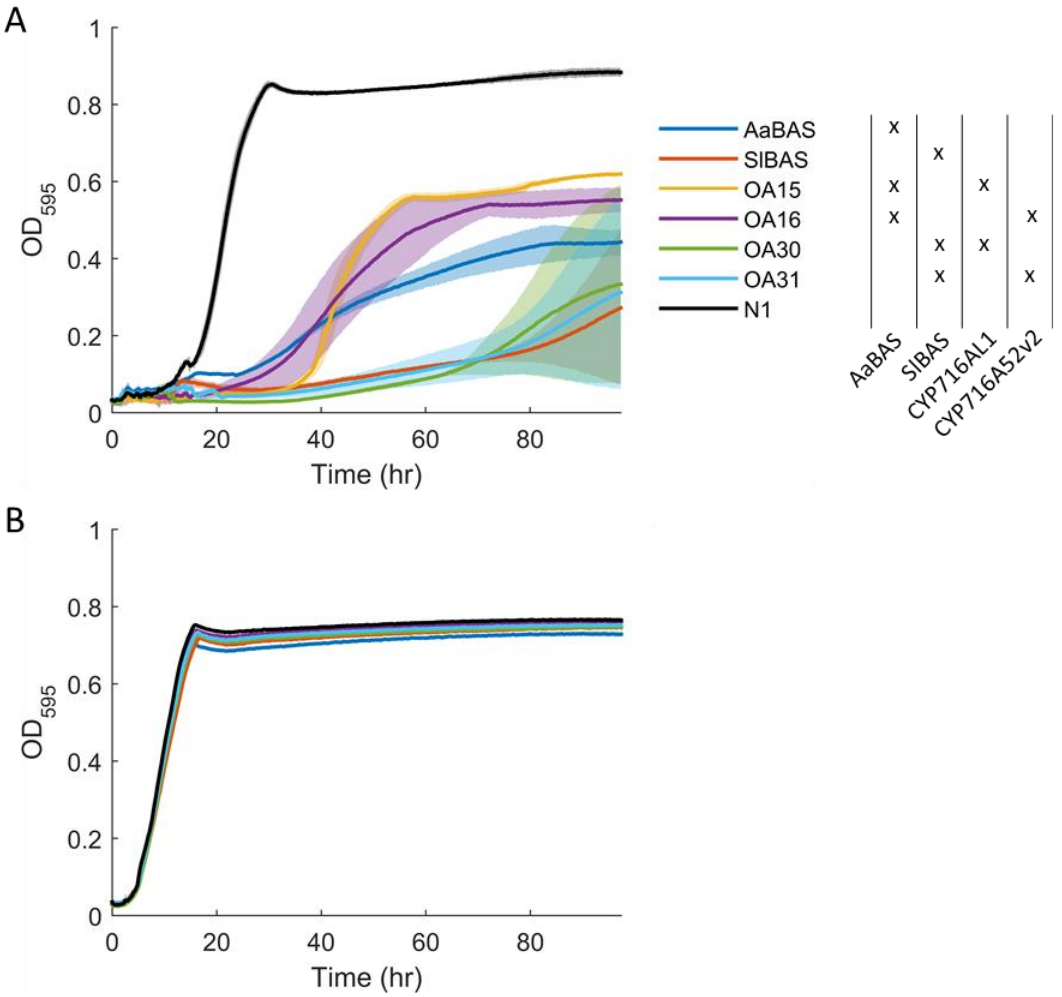


Figure 5.13. Growth of strains expressing "optimal" and "suboptimal" oleanolic acid pathways. The growth of strains expressing combinations of BASs and CYP716As with high (*AaBAS*, *CYP716AL1*) or low (*SlBAS*, *CYP716A52v2*) productivities were monitored using a 96-well plate assay. Cells were grown in galactose **(A)** or glucose **(B)** supplemented media and OD_{595} was measured every 15 min for 96 h in a Tecan Sunrise plate reader. Strains expressing *SlBAS* grew more slowly than those expressing *AaBAS*, indicating that the BAS homologue is the primary determinant of strain growth.

growth dynamics of MD-OA15 and 16 are genuinely different. Nevertheless, both display different growth curves to the AaBAS-only strain, indicating that the additional expression of the P450 and ATR2 affects growth. It is interesting that expressing the P450 and ATR2 did not have an observable effect on the growth rate of the SlBAS strain. This might be because the SlBAS-only strain already grew very slow, meaning that any additional effect of P450 and ATR2 expression could not be observed. As expected, MD-N1 grew considerably faster than the other strains in galactose-supplemented media (Figure 5.13 A). Together, the data suggest that while the additional expression of the P450 and ATR2 can affect growth, the choice of BAS is the primary determinant of growth rate.

Triterpenoid and ergosterol titres for the shake flask cultures were similar to those observed for the comparative analysis, which was performed in glass culture vials (Figure 5.18). One notable exception was the oleanolic acid titre of MD-OA15, which was 8.5 mg/L in the shake flask experiment compared with 14.1 mg/L in the culture vials. The cause of this is unknown, but could possibly be explained by differences in aeration (the P450-catalysed reaction is oxygen-dependent). The final OD_{600} was also similar for both shake flask and glass vial cultures. Thus, in general, culturing in shake flasks had a negligible impact on growth and productivity.

5.7 Comparing AaBAS and AsBAS for oleanolic acid production

The BAS comparative analysis showed that *AsBAS* expression had little effect on growth rate compared with an empty vector control strain (MD-N1), yet still produced a considerable amount of β-amyrin (Figure 5.8 A, Figure 5.7 C). By contrast, the strain expressing *AaBAS* grew much slower compared with *AsBAS* and MD-N1, and was used in the CYP716A comparative analysis due to its superior β-amyrin titres. The CYP716A strains grew much more slowly compared with MD-N1, probably due to burden caused by expression of the triterpenoid biosynthetic pathway, and it is possible that this may limit productivity. The use of *AsBAS* instead of *AaBAS* may afford a faster growth rate, which could affect triterpenoid titres. A strain expressing *AsBAS* and *CYP716AL1* (MD-OA29) was therefore analysed for triterpenoid production by GC-MS, culturing in shake flasks as above. Growth was also monitored using the microplate assay. MD-OA15 (*AaBAS*, *CYP716AL1*) and MD-N1, as well as

strains expressing *AaBAS* or *AsBAS* alone (MD-BA14 and MD-BA1), were also analysed as controls.

There was a considerable difference in triterpenoid titres and product ratios between the two strains (Figure 5.14 A-B). While MD-OA15 produced substantial amounts of β-amyrin, erythrodiol, oleanolic aldehyde and oleanolic acid, MD-OA29 accumulated predominantly oleanolic acid. β-amyrin and erythrodiol titres could not be calculated for MD-OA29 as the peak areas were too low, while the oleanolic acid titres were similar to MD-OA15 (7.1 mg/L for MD-OA29 and 8.5 mg/L for MD-OA15). MD-OA29 reached a slightly higher OD_{600} than MD-OA15 (7.8 vs. 6.1) and produced more ergosterol (Figure 5.14 C-D).

The growth curves showed that MD-OA15 and MD-OA29 also grew at considerably different rates in galactose-supplemented media, with MD-OA29 growing much faster than MD-OA15 (Figure 5.15 A). Interestingly, MD-OA29 grew at a similar rate to MD-BA1 (*AsBAS* alone), while MD-OA15 grew more similarly to MD-BA14 (*AaBAS* alone). The growth of MD-OA15 was a little different to MD-BA14, having a longer lag phase followed by a fast exponential, whereas MD-BA14 had a long exponential phase characterised by slow growth. Thus, the additional expression of *CYP716AL1* and *ATR2* had little effect on growth rate in the AsBAS strain, while it adjusted the growth dynamics in the AaBAS strain but still had much slower growth compared with MD-N1. As expected, all strains grew similarly in glucose-supplemented media, although the strains expressing BAS alone reached a slightly lower OD_{595} than the other strains (Figure 5.15 B).

The data thus suggest that much of the observed burden is caused by the BAS rather than the P450 and ATR2. This is consistent with the growth curves from the CYP716A comparative analysis, where none of the strains grew slower than MD-BA14 (Figure 5.8 A). It is also consistent with the growth curves in the previous section, where the oleanolic acid strains grew similarly to the strains expressing only the corresponding BAS (Figure 5.13 A).

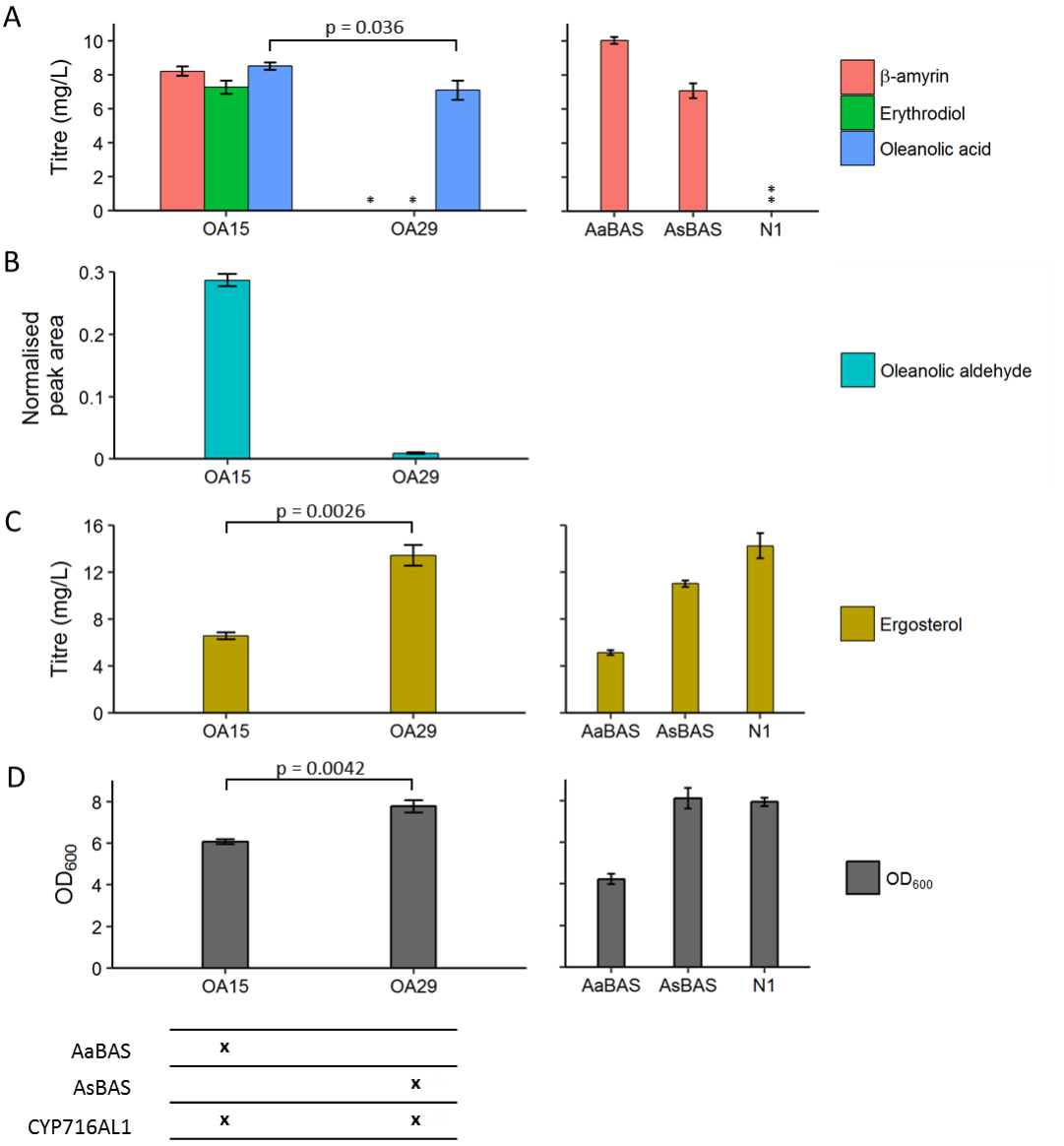


Figure 5.14. Comparison of strains expressing *AaBAS* and *AsBAS* for oleanolic acid production. Strains expressing either *AaBAS* or *AsBAS* in an oleanolic acid pathway (BAS, *CYP716AL1* and *ATR2*) were analysed for triterpenoid production by GC-MS. Control strains expressing the BAS alone or carrying an empty vector (MD-N1) were also analysed. **(A)** β-amyrin, erythrodiol and oleanolic acid titres. **(B)** Oleanolic aldehyde normalised peak areas. **(C)** Ergosterol titres. **(D)** OD_{600} at the end of the production run. MD-OA29, which expressed *AsBAS*, produced predominantly oleanolic acid, with little β-amyrin, erythrodiol and oleanolic aldehyde detected. By contrast, MD-OA15, which expressed *AaBAS*, produced a more even mixture of these triterpenoids. *, titre could not be calculated as the peak area fell below the linear region of the standard curve; **, no peak observed in the TIC. Shown is the mean ± 1 standard deviation for 3 biological replicates. P-values calculated using a Welch (unequal variances) two-sample t-test in R.

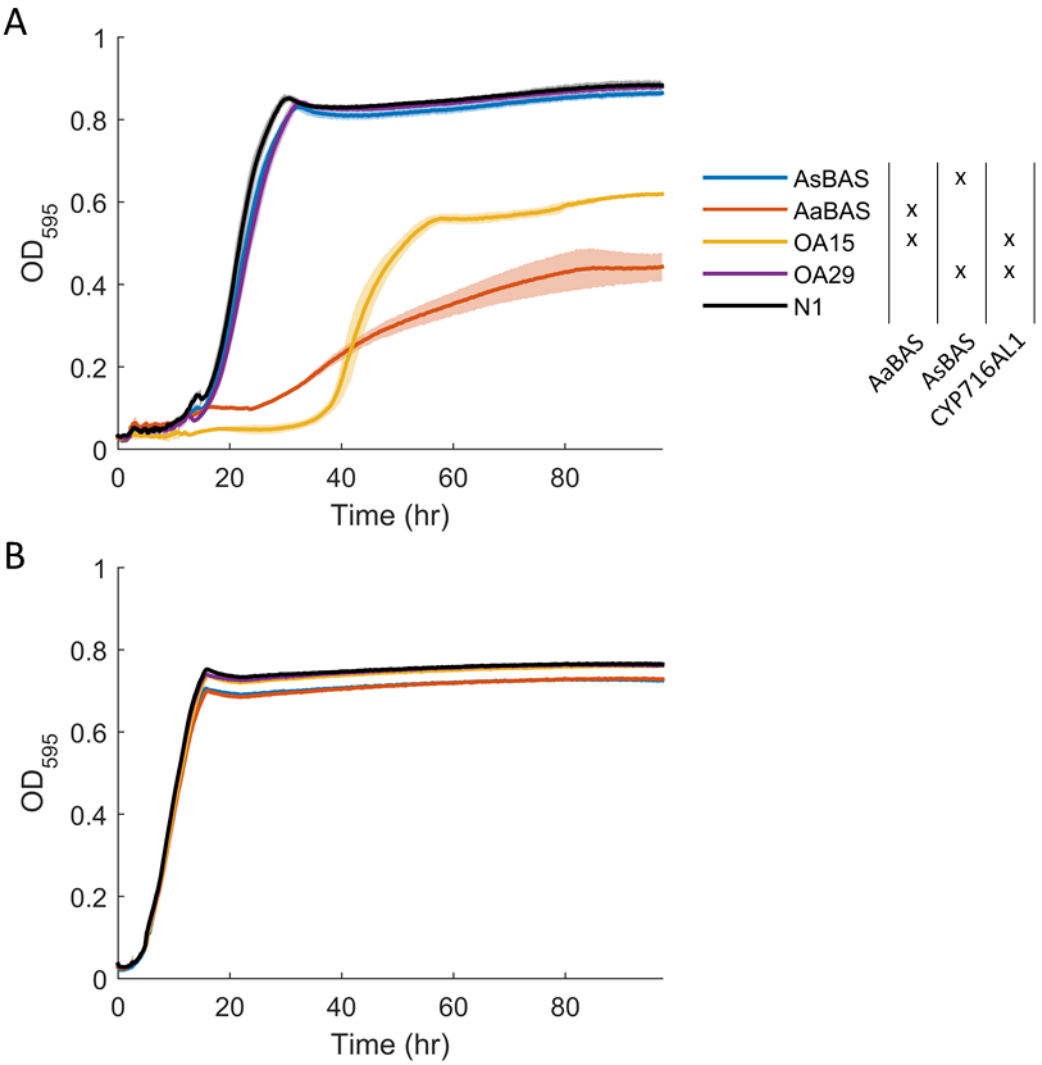


Figure 5.15. Growth of strains expressing oleanolic acid pathways containing either *AaBAS* or *AsBAS*. The growth of strains MD-OA15 (*AaBAS*, *CYP716AL1*, *ATR2*) and MD-OA29 (*AsBAS*, *CYP716AL1*, *ATR2*) were monitored using a 96-well plate assay. Strains expressing the BAS alone or carrying an empty vector (MD-N1) were also analysed. Cells were grown in galactose **(A)** or glucose **(B)** supplemented media and OD_{595} was measured every 15 min for 96 h in a Tecan Sunrise plate reader. MD-OA29 grew at similar rate to AsBAS and MD-N1 in galactose, while MD-OA15 and AaBAS grew much more slowly.

5.8 BioLector fermentations of oleanolic acid and β-amyrin strains

To assess strain growth during an actual fermentation run, cells were cultured using a BioLector micro fermentation system (m2p-labs GmbH, Baesweiler, Germany). This allows biomass to be measured throughout the fermentation run, and cultures can be harvested at the end of culturing for GC-MS analysis. During the 96-well plate assays performed earlier, it was noticed that cells had settled to the bottom of the wells by the end of the culture period, which may have affected strain growth. Indeed, plate readers generally have low shaking speeds and stop shaking during measurements, which contributes to the settling of cells. Furthermore, oxygen transfer is known to be poor when growing cells in 96-well plates due to the poor shaking regime, and the lack of humidity control can result in evaporation from the wells, which can have an impact on results (Kensy et al., 2009). The BioLector uses 48-well plates with baffled wells and continuous shaking that continues during measurements. This gives a similar agitation to that seen in conventional bioreactors and should prevent cells from settling and improve oxygen transfer (Kensy et al., 2009; Back et al., 2016). In addition, humidity can be maintained to prevent evaporation. The BioLector uses scattered light to measure cell biomass, which allows measurements of higher biomass concentrations to be made compared with OD_{600} readings, which are known to be inaccurate at high biomass concentrations (Back et al., 2016). Thus, growth curves generated using the BioLector should be more representative of strain growth during a fermentation run.

Strains MD-OA15 and MD-OA29, as well strains expressing only a BAS (either *AaBAS* or *AsBAS*) and MD-N1, were cultured in the BioLector for 96 h. Biomass was measured every 15 min, and the OD_{600} was measured and cellular metabolites were analysed by GC-MS at the end of the fermentation. The results were similar to the shake flask experiments above, with MD-OA29 producing mainly oleanolic acid, and reaching a higher OD_{600} and producing more ergosterol than MD-OA15 (Figure 5.16). However, MD-OA29 produced more oleanolic acid than MD-OA15, and the strain expressing only *AsBAS* produced a similar amount of β-amyrin to *AaBAS* (Figure 5.16 A). This contrasted with the results from the shake flask experiment, where MD-OA15

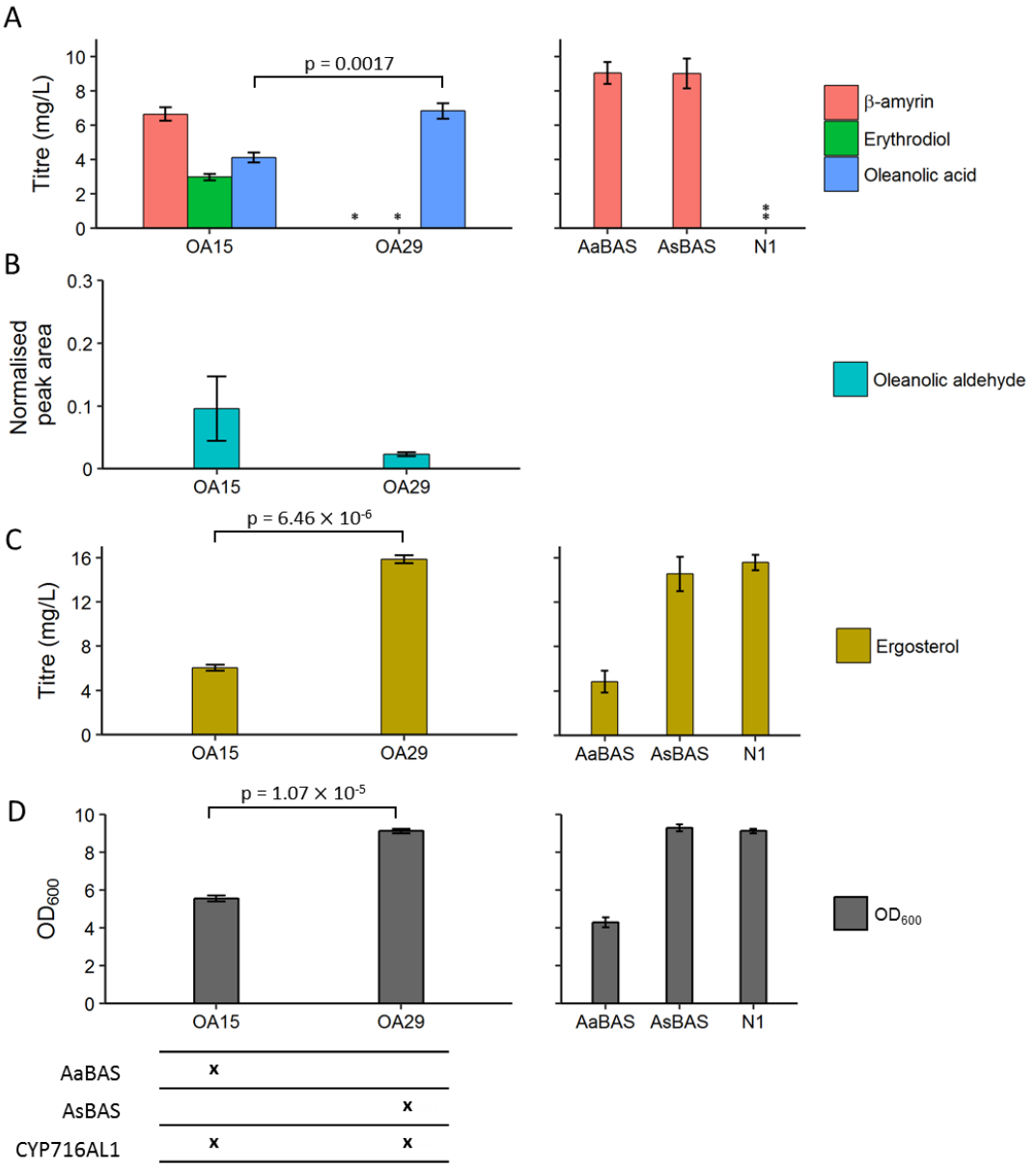


Figure 5.16. BioLector fermentations for oleanolic acid production. Strains MD-OA15 and MD-OA29, which expressed oleanolic acid pathways containing *AaBAS* or *AsBAS*, respectively, were grown in a BioLector for 96 h and triterpenoid production was measured by GC-MS. Control strains expressing the BAS alone or carrying an empty vector (MD-N1) were also analysed. **(A)** β-amyrin, erythrodiol and oleanolic acid titres. **(B)** Oleanolic aldehyde normalised peak areas. **(C)** Ergosterol titres. **(D)** OD_{600} at the end of the production run. *, titre could not be calculated as the peak area fell below the linear region of the standard curve; **, no peak observed in the TIC. Shown is the mean ± 1 standard deviation for 3 biological replicates. P-values calculated using a Welch (unequal variances) two-sample t-test in R.

produced more oleanolic acid than MD-OA29, and AaBAS gave higher β-amyrin titres than AsBAS (Figure 5.14 A).

Similar to the growth curves generated using the plate reader assay above, MD-OA15 grew similarly to the strain expressing *AaBAS* alone, while MD-OA29 grew at a similar rate to the strain expressing *AsBAS* alone (Figure 5.17, Figure 5.15 A). However, the shapes of the growth curves were a little different than in the plate reader assay. Notably, MD-OA15 had a shorter lag phase and a much longer and slower exponential phase in the BioLector fermentation, similar to AaBAS. By contrast, in the plate reader assay it had a long lag phase followed by a relatively fast exponential phase (Figure 5.15 A). This difference may be due to the improved agitation and oxygenation provided by the BioLector, which could allow cells to start growing sooner after inoculation, contributing to a shorter lag phase.

Furthermore, in the plate reader assay growth slowed considerably following the exponential phase and appeared to approach a stationary phase by the end of the culture (Figure 5.15 A). By contrast, growth in the BioLector slowed by a much smaller degree following the exponential phase and did not appear to be approaching a stationary phase by the end of the fermentation (Figure 5.17). Yeast growth is typically characterised by a lag phase, an exponential phase where fermentation and respiration of sugars occurs, a diauxic shift, a post-diauxic growth phase where cells perform only respiration, and finally a stationary phase (Teste et al., 2009). This was the pattern observed for MD-OA29, MD-N1 and AsBAS when grown in the BioLector, although it was less apparent for MD-OA15 and AaBAS, perhaps because of their slow growth (Figure 5.17). In the plate reader assay, the growth period following the exponential phase likely corresponds to this post-diauxic growth phase (Figure 5.15 A). The fact that growth appears to be much slower here compared with the BioLector fermentations could be due to poor mixing and oxygenation in the 96-well plate, or due to issues with measuring OD_{595} at high cell densities. The fact that MD-OA15, which reached a lower OD_{595} than MD-OA29, MD-N1 and AsBAS, also showed slow post-exponential growth in the plate reader assay indicates that the former is the case.

A comparison of selected strains grown in culture vials (from the CYP716A comparative analysis), shake flasks and the BioLector is presented in Figure 5.18. Titres and the final OD_{600} were generally similar across all three experiments. However, the oleanolic acid titre for MD-OA15 was only 8.5 mg/L in the shake flask

experiment and 4.1 mg/L for the BioLector fermentation, compared with 14.1 mg/L for the culture vials (Figure 5.18 D). Erythrodiol titres for MD-OA15 were similar between the culture vials and shake flasks (6.7 mg/L and 7.3 mg/L, respectively), but only 3.0 mg/L for the BioLector (Figure 5.18 B). It is unclear why oleanolic acid and erythrodiol titres for MD-OA15 were lower in the BioLector fermentation compared with the culture vials and shake flasks, especially considering that β-amyrin titres, ergosterol titres and the final OD_{600} were similar across all experiments for this strain, and that the other strains did not show such a difference.

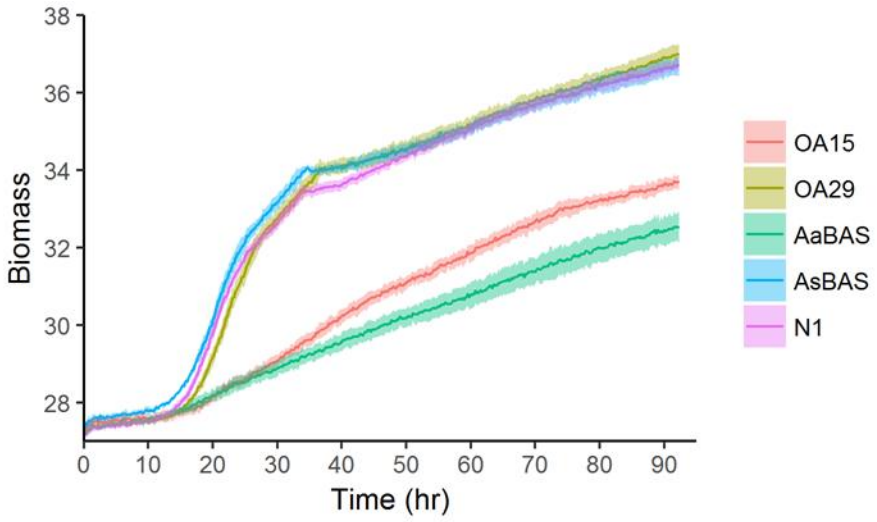


Figure 5.17. Growth of strains in the BioLector. Biomass was measured every 15 min for the duration of the fermentation (96 h) with continuous shaking and humidity maintained at 85 % relative humidity. MD-OA29, AsBAS and MD-N1 grew similarly, while MD-OA15 and AaBAS grew more slowly. This corroborates the data from the 96-well plate assays, suggesting that the choice of BAS is the most important determinant for growth. Shown is the mean ± 1 standard deviation for 3 biological replicates.

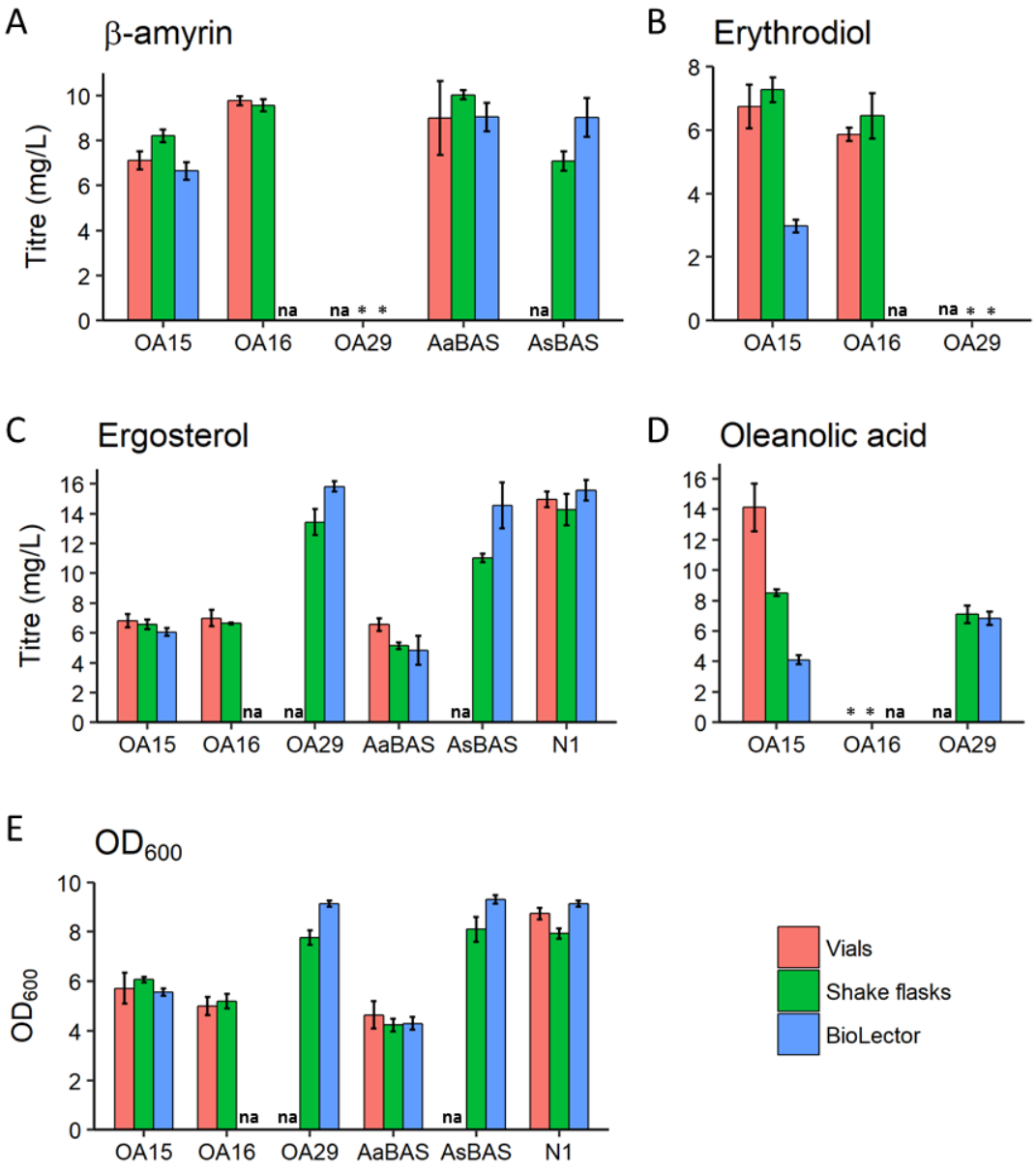


Figure 5.18. Comparison of culture vials, shake flasks and the BioLector for triterpenoid production. Titres of β-amyrin **(A)**, erythrodiol **(B)**, ergosterol **(C)** and oleanolic acid **(D)**, and the OD_{600} at the end of the production runs **(E)**, are shown. Data for the culture vials are from the CYP716A comparative analysis. Titres and OD_{600} were similar for all methods of culture, with the exception of MD-OA15 erythrodiol and oleanolic acid titres, which were lower in the BioLector experiment (erythrodiol) or in both the shake flask and BioLector experiments (oleanolic acid). *, titre could not be calculated as the peak area fell below the linear region of the standard curve; na, strain not analysed in this experiment. Shown is the mean ± 1 standard deviation for 3 biological replicates.

5.9 Discussion

Sixteen CYP716As were compared for their ability to oxidise β-amyrin to oleanolic acid when expressed in yeast. A range of activities and product ratios were observed, with calculated oleanolic acid titres ranging from 14.1 mg/L (for CYP716AL1) to 2.08 mg/L (for CYP716A179), a 6.8-fold difference (Figure 5.7 A). Interestingly, CYP716A75 produced 20.9 mg/L erythrodiol, representing the highest triterpenoid titre in the present study, and accumulated very little oleanolic aldehyde or oleanolic acid (Figure 5.7 A-B). By contrast, CYP716A110 produced only 1.12 mg/L erythrodiol, an 18.7-fold difference than CYP716A75. Growth assays showed that all strains generally grew similarly, although they could be divided into two groups based on the shape of their growth curves (Figure 5.10 A).

Notably, these titres were reported in strains that expressed only the oleanolic acid biosynthetic pathway and were otherwise unoptimised for triterpenoid production. Higher oleanolic acid titres have been reported in strains that were metabolically engineered to boost flux towards triterpenoid biosynthesis. Dai *et al.* reported an oleanolic acid titre of 71 mg/L in a strain expressing two copies of *GgBAS*, *CYP716A12*, *ATR1* and overexpressing the MEV and early sterol pathway genes *tHMG1*, *ERG9* and *ERG1*, which should increase flux towards 2,3-oxidosqualene (Dai et al., 2014). These genes were integrated and expressed from constitutive promoters, allowing the strains to be cultured in rich media containing glucose as the carbon source (YPD media). This should allow for improved growth and possibly higher triterpenoid production compared with the present study, which used a plasmid-based expression system that required the use of synthetic media with galactose as the carbon source, because glucose is the favoured carbon source for yeast (Timson, 2007).

The highest oleanolic acid titre reported to date was 186 mg/L (in shake flasks) and 607 mg/L (in a fermenter) (Zhao et al., 2018). This was achieved through several strategies. First, four CPRs were compared in a strain expressing *GgBAS* and *CYP716A12* (all expressed from the galactose-inducible *GAL* promoters), with *MtCPR* (*Medicago truncatula*) being identified as the optimal CPR, and this pathway was then integrated. Second, the *GAL1* and *GAL80* genes were knocked out. They are responsible for the metabolism of galactose and the repression of the *GAL* promoters in the absence of galactose, respectively, and their knockout was therefore

be expected to increase expression from the *GAL* promoters. Finally, flux towards 2,3-oxidosqualene was increased through overexpression of *tHMG1*, *ERG9* and *ERG1* (Zhao et al., 2018). Thus, there is great promise to scale-up oleanolic acid production towards industrially relevant levels.

A comparative analysis of P450s of the CYP93E subfamily, which catalyse the *C*-24 hydroxylation of β-amyrin, was previously reported by Moses *et al.* (Moses et al., 2014c). Similar to the results obtained in the present study, the CYP93Es showed a range of activities and product ratios. Together with the results presented here, this highlights the importance of screening homologues for metabolic engineering applications. The CYP716As were co-expressed with *AaBAS* and *ATR2* for the comparative analysis; AaBAS is one of the most productive BASs (together with CqBAS1) identified from the BAS comparative analysis (see Chapter 4). By comparing strains expressing combinations of genes with high (*AaBAS*, *CYP716AL1*) and low (*SlBAS*, *CYP716A52v2*) oleanolic acid productivities, it was shown that combining the most productive homologues resulted in considerably higher oleanolic acid production (Figure 5.12). MD-OA15, which expressed *AaBAS* and *CYP716AL1*, produced 8.5 mg/L oleanolic acid, while oleanolic acid titres could not be calculated for MD-OA31 (expressing *SlBAS* and *CYP716A52v2*) due to its low abundance.

Further to this, strains expressing *AaBAS* and *AsBAS* (co-expressed with *CYP716AL1* and *ATR2*) were compared for oleanolic acid production. In Chapter 4 it was observed that the strain expressing *AsBAS* grew much faster than the other BASs, having a very similar growth rate to the empty vector control strain, and this might be advantageous for oleanolic acid production. The strain expressing *AsBAS* (MD-OA29) produced nearly as much oleanolic acid as MD-OA15 (expressing *AaBAS*) (7.1 mg/L compared with 8.5 mg/L), yet accumulated minimal quantities of β-amyrin, erythrodiol and oleanolic aldehyde (titres undetermined) (Figure 5.14). This further highlights the impact that the choice of homologue has on strain performance.

It is unclear why MD-OA29 accumulated mainly oleanolic acid, but it is possible that the lower burden caused by *AsBAS* expression is responsible. Comparison with characterised OSCs (notably human and yeast lanosterol synthases) suggests that BASs may be monotopic membrane proteins localised to the ER or lipid particles (which are closely associated with the ER) (Milla et al., 2002; Thoma et al., 2004; Thimmappa et al., 2014; Jacquier et al., 2013; Liang et al., 2012). Furthermore, most

plant P450s and CPRs localise to the ER via an N-terminal membrane anchor, and may also localise here in yeast (Bak et al., 2011; Renault et al., 2014; Urban et al., 1994). Overexpression of the BASs from the strong *GAL1* promoter may saturate the ER, which could disrupt CYP716A activity. Western blots were unable to detect AsBAS protein (Figure 4.11 A), possibly indicating that it is expressed at a lower level compared to the other BASs. However, further work to detect AsBAS and compare its protein level with the other BASs would be required to confirm this.

An interesting observation from the growth assays performed in conjunction with the above experiments was that strain growth appeared to largely depend on the choice of BAS. The additional expression of a CYP716A and *ATR2* had little impact on growth rate compared with a strain expressing only a BAS (Figure 5.10 A, Figure 5.13 A, Figure 5.15 A). Furthermore, the choice of CYP716A had only a small effect on growth rate, while changing the BAS altered growth considerably. Specifically, MD-OA15 and MD-OA16 (expressing *AaBAS*, *ATR2* and either *CYP716AL1* or *CYP716A52v2*) grew similarly to the strain expressing only *AaBAS* (Figure 5.13 A); MD-OA30 and MD-OA31 (expressing *SlBAS*, *ATR2* and *CYP716AL1* or *CYP716A52v2*) grew similarly to the strain expressing only *SlBAS* (Figure 5.13 A); and MD-OA29 (expressing *AsBAS*, *ATR2* and *CYP716AL1*) grew similarly to the strain expressing only *AsBAS* (Figure 5.15 A). Thus, BAS expression appears to be the primary cause of burden.

To investigate the causes of this burden further, it will be essential to determine the expression levels of the BASs, CYP716As and *ATR2*. Similar to what was discussed in Chapter 4, variation in expression could explain the different catalytic activities of the CYP716As and the differences in growth observed when expressing different BASs. Western blots of four CYP716As and ATR2 were unable to detect any protein (Figure 5.11 A), while several BASs (except for AsBAS) could be detected using this method (see Chapter 4, Figure 4.11 A). This might suggest that the BASs were more highly expressed than the CYP716As and *ATR2*, which could explain why the additional expression of a CYP716A and *ATR2* does not appear to substantially increase burden. However, quantitative information cannot be obtained from these western blots, which were preliminary experiments performed without replicates or loading controls. Further work to detect the CYP716As and ATR2 and compare their expression levels with the BASs is therefore required to confirm this hypothesis.

To determine expression levels therefore, the western blots could be repeated and loading controls included. However, many replicates might be needed due to the inherent error in quantitative western blotting (Hoshino, 2017). As discussed in Chapter 4, the use of untargeted proteomics may be a better approach for comparing relative expression levels and may allow the detection of poorly expressed proteins. Indeed, a preliminary experiment was able to detect both CYP716AL1 and ATR2, as well as AaBAS, in MD-OA15 (Table 5.3). Although replicates were not included and quantification was not performed, the data indicate that AaBAS may indeed be more abundant than CYP716AL1 and ATR2, based on the number of peptide matches. Although this metric does not represent protein abundance, it does have a rough correspondence with abundance. AaBAS had 24 peptide matches, while CYP716AL1 and ATR2 had 7 and 5, respectively. Such large differences typically indicate differences in abundance, with a greater number of matches corresponding to higher abundance. Nevertheless, additional replicates will be required to perform quantification and thus obtain an accurate measure of relative expression levels.

In addition to determining expression levels, it would be interesting to investigate the burden caused by the expression of the CYP716As and *ATR2* without BAS expression. For example, the growth of strains expressing a CYP716A and *ATR2* but no BAS, a CYP716A only, or *ATR2* only could be monitored. This would allow the burden caused by CYP716A and *ATR2* expression to be investigated in the absence of a BAS, to better understand the individual contributions of these enzymes to the observed burden.

Although differences in expression may explain variation in the total catalytic activity of the CYP716As, they should not influence the product profiles. Instead, variations in product profiles are most likely caused by differences in the sequences of the CYP716As. Although *in silico* analyses of the CYP716A sequences were performed, there were no apparent amino acid variants that might explain the observed differences in product profiles. Indeed, little is known about the catalytic mechanism of the CYP716A subfamily beyond the current understanding of P450 catalysis broadly. P450 substrate binding and recognition regions are highly variable (Werck-Reichhart and Feyereisen, 2000), which is not surprising given the diverse substrates utilised by the P450 superfamily. Thus, subfamily-specific knowledge is required to explain the substrate and product specificities of individual P450s. Nevertheless, the

use of homology and substrate docking models may allow the differences in product profiles to be probed further.

In addition to using culture vials, fermentations were performed in shake flasks and a BioLector micro fermenter. However, these failed to boost production, with titres generally being similar for all three experiments (Figure 5.18). Interestingly, erythrodiol titres for MD-OA15 were lower for the BioLector fermentations than for the culture vials and shake flasks. Furthermore, oleanolic acid titres for this strain were lower in both shake flasks and the BioLector than the culture vials. The reason for this is unclear, although it could be explained by differences in aeration. The CYP716As require molecular oxygen to oxygenate their substrates, and it is possible that inadequate aeration leading to lower intracellular oxygen availability could affect the oxygenation of β-amyrin by these enzymes, altering the product profile. However, as discussed earlier, the BioLector provides a high level of agitation and aeration to the cells, which would not be expected to reduce erythrodiol and oleanolic acid titres.

Table 5.3. Preliminary proteomics analysis of MD-OA15

Protein	Peptide matches (unique matches)	Peptides
AaBAS(Sc)	24 (13)	VDFWNHR
		HEVKPSSDVLWR
		HEVKPSSDVLWR
		IPDYLWVAEDGMK
		IPDYLWVAEDGMK
		IPDYLWVAEDGMK
		WILDHGSVTTIPSWGK
		WILDHGSVTTIPSWGK
		DELYAQPYDEIKWR
		KEIENFLLGSSGYLEK
		KEIENFLLGSSGYLEK
		CCLLFATMPPEIVGEK
		QIWEFDPNYGTPEER
		QIWEFDPNYGTPEER
		SNLVHTAWAMMGLIHSR
		KWILDHGSVTTIPSWGK
		KWILDHGSVTTIPSWGK
		KWILDHGSVTTIPSWGK
		KKEIENFLLGSSGYLEK
		KKEIENFLLGSSGYLEK
		FLLETQLEDGGWGESYK
		FLLETQLEDGGWGESYK
		GSWTFSDQDHGWQVSDCTAEALK
		GSWTFSDQDHGWQVSDCTAEALK
CYP716AL1(Sc)	7 (6)	STLPGPLPPGR
		YIGLMDQIAQK
		YSWNVACEVLR
		MLPNFLKPEALQR
		MLPNFLKPEALQR
		IVVDPMPIPEKGLPVR
		SHLLGEPAAVFCGAIGNK
AtATR2(Sc)	5 (5)	SENCSSAPIFVR
		VEPLKPLVIKPR
		MDFIYEEELQR
		IVDLDDYAADDDEYEEK
		NAVPYEKSENCSSAPIFVR

Chapter 6: Production of Oleanane Saponins

6.1 Introduction

In addition to oxygenations, triterpenoids are often subject to glycosylations catalysed by UDP-glycosyltransferases (UGTs) (Thimmappa et al., 2014). The resulting product is called a saponin, and glycosylations contribute to the diverse bioactivities found in this family of metabolites (Augustin et al., 2011; Thimmappa et al., 2014). Triterpenoid aglycones are typically hydrophobic, and the addition of a large polar moiety (the sugar) to the aglycone increases polarity and influences solubility (Thimmappa et al., 2014; Augustin et al., 2011). Glycosylation thus has a dramatic effect on the physicochemical properties of the triterpenoid, and is responsible for many of their activities, such as the ability to act as surfactants and foaming agents and the insecticidal activities of some saponins (Moses et al., 2014a; Augustin et al., 2012; Nielsen et al., 2010). Indeed, many saponins are believed to form pores in cell membranes due to their amphipathic structures, resulting in their haemolytic activities (Augustin et al., 2011; Moses et al., 2014a).

Glycosylations occur at oxygenated positions of the triterpenoid, most commonly at the *C*-3 hydroxyl group and the *C*-28 carboxyl group (Thimmappa et al., 2014). Most triterpenoids possess the *C*-3 hydroxyl, which is derived from the epoxide of 2,3-oxidosqualene, while a cytochrome P450 with *C*-28 oxidase activity is required to form the *C*-28 carboxyl (see Chapter 5). Triterpenoids can be glycosylated at multiple positions with different sugars, and the sugar chains can vary in length and be linear or branched (Moses et al., 2014a). The UGTs involved in saponin biosynthesis generally have high regiospecificities but a broader substrate range, glycosylating a number of structurally related triterpenoid substrates and sometimes non-triterpenoid substrates (Meesapyodsuk et al., 2007; Naoumkina et al., 2010; Achnine et al., 2005; Thimmappa et al., 2014; Augustin et al., 2012; Erthmann et al., 2018). Most saponins are monodesmosidic, having a single sugar chain, and many bidesmosidic saponins are also known, while saponins that are glycosylated at more than two positions are rare (Augustin et al., 2011). Sugar chain lengths typically range from two to five sugars, and common sugars include glucose, galactose and glucuronic acid, although many different sugars are possible (Augustin et al., 2011). Sugar chains are built up

by the sequential addition of monosaccharides to the triterpenoid, rather than the addition of a previously synthesised oligosaccharide (Augustin et al., 2011). The sugars can also be subject to further modifications by tailoring enzymes, such as acylations (Ragupathi et al., 2011).

The combined action of P450s (catalysing oxygenations), UGTs (catalysing glycosylations) and various tailoring enzymes (e.g. catalysing acylations) therefore allows plants to produce highly complex saponins (Moses et al., 2014a). For example, QS-21, a mixture of two saponins derived from a *Quillaja saponaria* extract, has a branched trisaccharide chain (glucuronic acid, xylose, galactose) at the *C*-3 hydroxyl and a linear tetrasaccharide chain at the *C*-28 carboxyl (fucose, rhamnose, xylose and either apiose or xylose). This tetrasaccharide is acylated at the fucose moiety, and the acyl group is itself glycosylated with an arabinose (Ragupathi et al., 2011). Plants also produce many simple saponins such as 3-*O*-Glc-oleanolic acid, which has a single glucose moiety at the *C*-3 hydroxyl (Augustin et al., 2012).

Many saponins have medically or industrially useful properties (Moses et al., 2014a). For example, glycyrrhizin is a sweetener produced by *Glycyrrhiza glabra* (liquorice) that has emulsifying and anti-viral activities (Ma et al., 2018; Fiore et al., 2008). Meanwhile, hederacolchiside A (*Pulsatilla koreana*) showed a stronger antitumour activity than the anticancer drugs doxorubicin and taxol in a mouse model (Bang et al., 2005). Quil A, a saponin extract from *Quillaja saponaria*, has been used in veterinary vaccines (Sun et al., 2009) while QS-21, a fraction derived from this extract, shows immunostimulatory activity (Moses et al., 2014a). QS-21 is used as an adjuvant in a vaccine against feline leukaemia virus (Sun et al., 2009) and also served an adjuvant in the anti-malarial vaccine RTS,S, although this vaccine showed low efficacy in a phase III clinical trial (Gosling and von Seidlein, 2016; Mahmoudi and Keshavarz, 2017). In addition, numerous saponins have been shown to have insecticidal or anti-feedant activities. For example, saponins derived from oleanolic acid and hederagenin acted as feeding deterrents against the plant pests *Phyllotreta nemorum* (flea beetle) and *Plutella xylostella* (diamondback moth), and this activity was dependent on the sugar moieties (Shinoda et al., 2002; Nielsen et al., 2010; Augustin et al., 2012). For more information, see section 1.2 and Figure 1.6.

Due to their diverse bioactivities it is desirable to develop microbial platforms for saponin production. However, only a limited range of saponins have been produced

in yeast, with most studies focusing on the production of triterpenoid aglycones. Only two β-amyrin-derived saponins (oleanane saponins) have been produced in yeast to date, and both were monoglucosides (Moses et al., 2014b; Arendt et al., 2017). 3-*O*-Glc-echinocystic acid is derived from β-amyrin by two P450-catalysed oxygenations and a single glycosylation at the *C*-3 hydroxyl (Moses et al., 2014b), while the production of 28-*O*-Glc-medicagenic acid involved three oxygenations and a glycosylation at the *C*-28 carboxyl (Arendt et al., 2017). Although oleanane saponins containing multiple sugars have not yet been produced in yeast, production of the diglycoside saponin Ginsenoside Rg3, which is derived from the tetracyclic triterpene dammarenediol-II, was recently demonstrated (Wang et al., 2015). Ginsenoside Rg3 has a Glc-Glc disaccharide chain at its *C*-3 hydroxyl, and its biosynthesis involves a dammarenediol synthase, a P450 and two UGTs. Nevertheless, as discussed above, many saponins that occur in nature are far more complex, having undergone many oxygenations, glycosylations and other modifications. Therefore, to access the diversity and realise the commercial potential of saponins, platforms that can produce such complex saponins must be developed.

Here, various UGTs (Table 6.1) were expressed in yeast strains engineered to produce β-amyrin or oleanolic acid, and saponin production was monitored by TLC and LC-MS. β-amyrin contains only the *C*-3 hydroxyl group, while oleanolic acid also possess the *C*-28 carboxyl group. UGTs expected to transfer a glucose moiety to these groups (UGT73C10 and UGT73F3) were tested (Figure 6.1), as well as two UGTs known to glycosylate triterpenoids but whose regiospecificities were unknown (UGT71G1 and UGT73K1) (Naoumkina et al., 2010; Augustin et al., 2012; Achnine et al., 2005). In addition, the co-expression of UGT73C10 and UGT73P2 was tested for the production of monodesmosidic saponin diglycosides (i.e. containing a disaccharide sugar chain at a single position). UGT73C10 adds a glucose to the *C*-3 hydroxyl, while UGT73P2 adds a galactose to the *C*-3 glucuronic acid of soyasapogenol B monoglucuronide in the biosynthesis of soyasaponin I (Shibuya et al., 2010; Augustin et al., 2012), and has also been shown to add a galactose to the *C*-3 glucose of 3-*O*-Glc-oleanolic acid (Figure 6.1) (Thomas Louveau, personal communication).

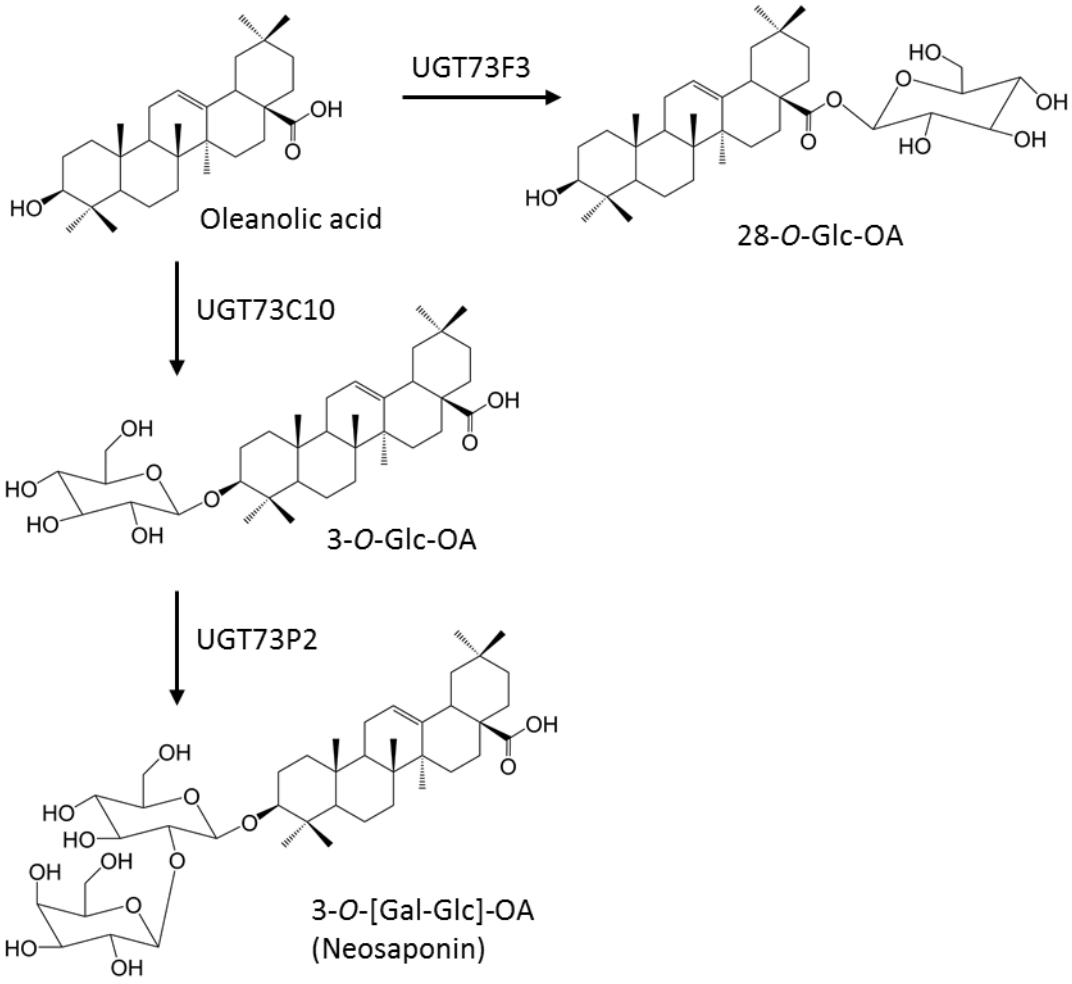


Figure 6.1. UGT-catalysed glycoslyations of triterpenoids. The glycosylation of oleanolic acid by UGT73C10, UGT73F3 and UGT73P2 are shown. UGT73C10 and UGT73F3 are expected to transfer a glucose moiety to the *C*-3 hydroxyl and *C*-28 carboxyl, respectively. UGT73P2 is expected to transfer a galactose to the glucose of 3-*O*-Glc-oleanolic acid, producing neosaponin. OA: oleanolic acid.

Table 6.1. UGTs tested in this study

UGT	Activity*	Species	Accession number	Reference
UGT73C10	*C*-3, Glc	*Barbarea vulgaris*	AFN26666	Augustin *et al.*, 2012
UGT73K1	Glc	*Medicago truncatula*	AAW56091	Achnine *et al.*, 2005
UGT71G1	Glc	*M. truncatula*	AAW56092	Achnine *et al.*, 2005
UGT73P2	*C*-2', Gal	*Glycine max*	BAI99584	Shibuya *et al.*, 2010
UGT73F3	*C*-28, Glc	*M. truncatula*	ACT34898	Naoumkina *et al.*, 2010

*The position of glycosylation and the sugar added are shown. Glc: glucose; Gal: galactose; *C*-2': carbon-2 of glucose.

6.2 TLC indicates strains are producing saponins

To generate yeast strains capable of producing saponins, strains MD-BA5 (expressing *EtBAS*) and MD-OA7 (expressing *EtBAS*, *CYP716A12* and *ATR2*) were transformed with plasmids expressing various UGTs (Table 6.1). UGT73C10 is a promiscuous UGT that can glycosylate a variety of triterpenoids at the *C*-3 hydroxyl group (Augustin et al., 2012). UGT73K1 and UGT71G1 were shown to glycosylate certain triterpenoids *in vitro* but their regiospecificities were not determined; it appeared likely that they glycosylated at the *C*-3 hydroxyl or *C*-28 carboxyl groups (Achnine et al., 2005). UGT73P2 can add a galactose to an existing glucose at the *C*-3 position (such as that added by UGT73C10), generating a disaccharide sugar chain (Thomas Louveau, personal communication). To date, the only oleanane saponins that have been produced in yeast were monoglycosides, containing a single sugar. Functional expression of *UGT73C10* and *UGT73P2* in yeast would thus allow the first production of a diglycoside saponin in yeast. It was shown that co-expression of *UGT73C10* and *UGT73P2* with genes for oleanolic acid biosynthesis in the tobacco plant *Nicotiana benthamiana* allows production of a novel saponin called neosaponin, which consists of an oleanolic acid aglycone with a Gal-Glc disaccharide at the *C*-3 hydroxyl (Thomas Louveau, personal communication). Neosaponin is so-called because it has not been observed in nature before, its production being enabled by the expression of a novel combination of enzymes in a heterologous host.

To monitor saponin production, yeast expressing the saponin biosynthetic pathways were grown for 90 h in galactose-supplemented medium. Butanol extractions were performed on both cells and the spent medium, and the extracts were analysed by TLC (Figure 6.2). MD-BA5, MD-OA7 and the empty vector control strain MD-N2 were also analysed, as were 0.6 µg of authentic β-amyrin, oleanolic acid and 3-*O*-Glc-echinocystic acid standards. 3-*O*-Glc-echinocystic acid is a saponin that is structurally similar to 3-*O*-Glc-oleanolic acid, having an additional hydroxyl group at *C*-16, and so should serve as a suitable control; standards for the saponins potentially produced by the strains in the present work were not available. The metabolite extraction and TLC methods were based on protocols provided by Thomas Louveau (Anne Osbourn lab, John Innes Centre). Bands were not observed for the β-amyrin and oleanolic acid standards, indicating that any bands observed in the samples are unlikely to correspond to triterpenoid aglycones (Figure 6.2). As expected, a band was observed for 3-*O*-Glc-echinocystic acid.

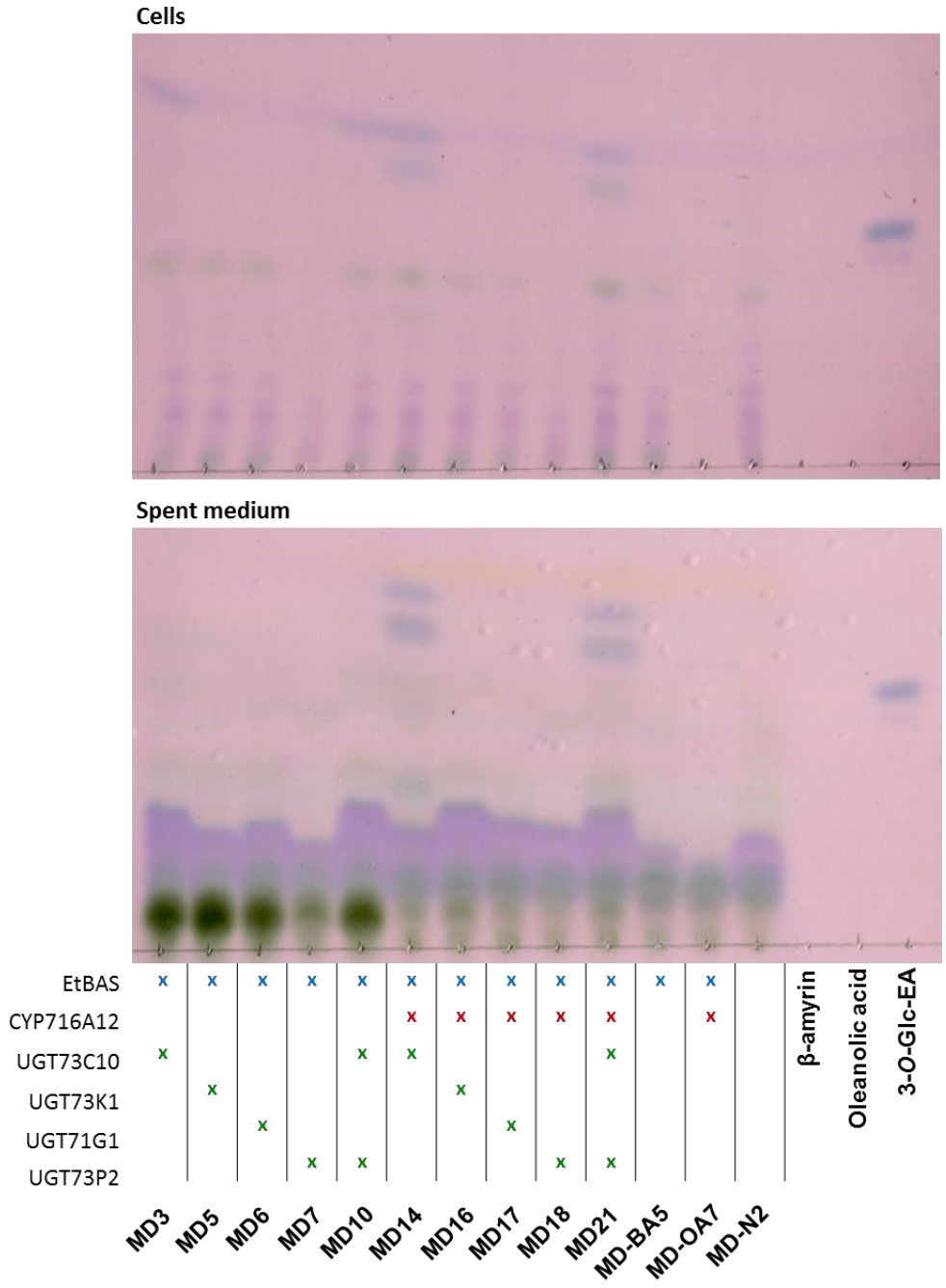


Figure 6.2. Thin layer chromatography (TLC) indicates saponin production. TLC was performed on extracts of cells co-expressing various UGTs with either *EtBAS* (for β-amyrin production) or an oleanolic acid pathway (*EtBAS*, *CYP716A12* and *ATR2*). Strains expressing *EtBAS* alone (MD-BA5), the oleanolic pathway alone (MD-OA7) or carrying empty vectors (MD-N2) were analysed as controls. Authentic standards of β-amyrin, oleanolic acid and the saponin 3-*O*-Glc-echinocystic acid (0.6 µg each) were also analysed. Top: cell extracts. Bottom: spent media extracts. The enzymes expressed by each strain are indicated.

Strain MD3, expressing *EtBAS* and *UGT73C10*, had a unique band that was not present for MD-BA5 (expressing only *EtBAS*) that may correspond to 3-*O*-Glc-β-amyrin. This band was only present in the cell extract (Figure 6.2). MD10 expressed EtBAS, *UGT73C10* and *UGT73P2*, so was expected to produce both 3-*O*-Glc-β-amyrin and 3-*O*-[Gal-Glc]-β-amyrin. However, only a single unique band, similar to MD3 and visible only in the cell extract, was observed. MD10 may therefore only produce 3-*O*-Glc-β-amyrin.

Strain MD14, expressing an oleanolic acid pathway (*EtBAS*, *CYP716A12* and *ATR2*) and *UGT73C10*, was expected to produce glycosylated derivatives of β-amyrin, oleanolic acid and the intermediates erythrodiol and oleanolic aldehyde. This strain had two unique bands, one that migrated a similar distance as that in the MD3 and MD10 cell extracts, and the other not present in those strains (Figure 6.2). Therefore, the first band may correspond to 3-*O*-Glc-β-amyrin, while the other may correspond to glycosylated erythrodiol, oleanolic aldehyde or oleanolic acid. The additional expression of *UGT73P2* (strain MD21) was expected to produce neosaponin (3-*O*-[Gal-Glc]-oleanolic acid) and possibly other saponin diglycosides derived from β-amyrin, erythrodiol or oleanolic aldehyde. However, this strain appeared identical to MD14, indicating that diglycosides either were not present or were not detected.

Both unique bands in strains MD14 and MD21, including that putatively belonging to 3-*O*-Glc-β-amyrin, were present in the cell and spent medium extracts. Interestingly, this contrasted with strains MD3 and MD10, where the putative 3-*O*-Glc-β-amyrin band was observed only in the cell extract (Figure 6.2). It is possible that production of glycosylated derivatives of oleanolic acid (or the intermediates erythrodiol or oleanolic aldehyde) results in secretion of 3-*O*-Glc-β-amyrin into the growth medium, for example by cell lysis caused by production of these saponins. Indeed, some oleanane saponins are reported to have haemolytic properties, lysing red blood cells, and saponins are believed to be capable of permeabilising cells membranes (Augustin et al., 2011). Furthermore, oleanane saponins produced by *Medicago truncatula* are classified as haemolytic and non-haemolytic saponins, and the haemolytic saponins all contain a *C*-28 carboxyl group, while the non-haemolytic saponins do not possess this moiety (Moses et al., 2014a). Alternatively, it is possible that one or more of the additional saponins potentially produced by MD14 and MD21 (i.e. derived from erythrodiol, oleanolic aldehyde and/or oleanolic acid) have the same

retention time as 3-*O*-Glc-β-amyrin, and it is these more oxidised saponins that are being secreted into the growth medium, rather than 3-*O*-Glc-β-amyrin. To test these hypotheses, it will be necessary to perform LC-MS to enable the identification of each saponin.

No products were detected in the strains expressing *UGT73K1* or *UGT71G1* (strains MD5, MD6, MD16 and MD17) (Figure 6.2), indicating that these UGTs may not be able to glycosylate β-amyrin or its *C*-28 oxygenated derivatives. These UGTs were previously shown to glycosylate certain triterpenoids derived from oleanolic acid in an *in vitro* assay but were not tested with other triterpenoids (Achnine et al., 2005). It is also possible that they did have activity against these substrates but the products were not abundant enough to be detected by TLC. Finally, strains MD7 (expressing *EtBAS* and *UGT73P2*) and MD18 (expressing the oleanolic acid pathway and *UGT73P2*) did not produce any detectable products. This was expected, because *UGT73P2* is reported to only extend an existing sugar chain at the *C*-3 hydroxyl, and is not able to glycosylate a triterpenoid aglycone (Shibuya et al., 2010).

6.3 LC-MS confirms saponin production and allows identification of individual saponins

The TLC results indicated that some strains produced saponins, however, it did not provide information on their identities. Therefore, LC-MS was performed to identify which saponins were produced. In addition to confirming the identity of metabolites through analysis of their mass spectra, the high sensitivity of this technique should also allow the detection of low abundance saponins that may not have been detected by TLC.

The current study used a PFP reverse phase column with a methanol gradient (i.e. a hydrophobic stationary phase and polar mobile phase). Hydrophobic analytes should thus elute later as they interact more strongly with the column, and saponins that have undergone more glycosylations or oxygenations should elute earlier than those that have not. Thus neosaponin (3-*O*-[Gal-Glc]-oleanolic acid), which has two sugars, should elute earlier than 3-*O*-Glc-oleanolic acid, while 3-*O*-Glc-β-amyrin should elute even later, as *C*-28 is a methyl group instead of a carboxyl like oleanolic acid. The soft ionisation technique (electrospray ionisation; ESI) used by the current LC-MS system results in less fragmentation of the analyte compared with the hard ionisation techniques employed in the GC-MS assays. As such, saponins that are analysed by

LC-MS do not undergo retro Diels-Alder (rDA) fragmentations, and the main ions observed should be the parent ion and fragment ions resulting from the loss of sugars and carboxyl groups.

In addition to the above strains analysed by TLC, another strain generated by the transformation of MD-OA7 with a plasmid containing *UGT73F3* was tested. UGT73F3 was reported to transfer a glucose to the *C*-28 carboxyl of the oleanolic acid derivatives hederagenin and medicagenic acid (Naoumkina et al., 2010; Arendt et al., 2017), and may therefore glycosylate oleanolic acid at this position. Strains were cultured and metabolites extracted as before, and the cell and spent media extracts were analysed by LC-MS (positive ion mode). Hannah Florance (University of Edinburgh) provided help in developing the LC-MS protocol. Extracted ion chromatograms (EICs) were generated for each expected saponin. EICs differ from total ion chromatograms (TICs) in that the intensity of only the specified ion is plotted in an EIC, whereas the total intensity of all ions are plotted in a TIC. Because they plot the intensity of only a single ion, EICs can be used to identify compounds whose peaks are otherwise hidden by noise. For each expected saponin, the sodium adduct of the parent ion was used to generate EICs. A summary of the results is shown in Table 6.2 and Table 6.3.

Although mass spectra were analysed, only the parent ion could be observed, indicating that the saponins underwent little or no fragmentation using the current method (Figure 6.3). The mass spectra therefore provide little additional information over the EICs. The lack of observed fragmentation therefore means that saponins cannot be absolutely identified, as less information is available for each observed peak. Nevertheless, the LC-MS method still provides additional information over the TLC in the form of the parent ion.

Table 6.2. Summary of LC-MS analyses (cell extracts)

Strain	Enzymes	m/z 611 (3-*O*-Glc-β-amyrin)	m/z 627 (3/28-*O*-Glc-erythrodiol)	m/z 625 (3-*O*-Glc-OAld)	m/z 641 (3/28-*O*-Glc-OA)	m/z 773 (3-*O*-[Gal-Glc] -β-amyrin)	m/z 789 (3-*O*-[Gal-Glc] -erythrodiol)	m/z 787 (3-*O*-[Gal-Glc] -OAld)	m/z 803 (3-*O*-[Gal-Glc] -OA)
MD3	UGT73C10	13.4’	-	-	-	-	-	-	-
MD5	UGT71G1	-	-	-	-	-	-	-	-
MD6	UGT73K1	-	-	-	-	-	-	-	-
MD10	UGT73C10 + UGT73P2	13.4’	-	-	-	-	-	-	-
MD14	CYP716A12 + UGT73C10	13.4’	10.8’ (3) 10.9’ (28?)	11.8’	10.9’ (3)	-	9.9’ (Glc-Glc)	-	9.6’ (Glc-Glc)
MD15	CYP716A12 + UGT73F3	-	10.9’ (28?)	-	10.6’ (28)	-	9.9’ (Glc-Glc)	-	9.6’ (Glc-Glc)
MD16	CYP716A12 + UGT71G1	-	-	-	-	-	-	-	-
MD17	CYP716A12 + UGT73K1	-	-	-	-	-	-	-	-
MD21	CYP716A12 + UGT73C10 + UGT73P2	13.4’	10.8’ (3) 10.9’ (28?)	11.8’	10.9’ (3)	-	9.9’ (Glc-Glc) 10.5’ (Gal-Glc)	-	9.6’ (2x Glc) 10.5’ (Gal-Glc)

All strains expressed *EtBAS*. The m/z value refers to the sodium adduct of the parent ion, $[M+Na]^+$, of the corresponding saponin. Retention times (min) of the observed peaks are shown. The potential position of the sugar (3 or 28) or the hypothesised sugar chain (Glc-Glc, Gal-Glc) are shown in brackets. OAld: oleanolic aldehyde; OA: oleanolic acid; Glc: glucose; Gal: galactose; 3-*O*-[Gal-Glc]-OA: neosaponin.

Table 6.3. Summary of LC-MS analyses (spent media)

Strain	Enzymes	m/z 611 (3-*O*-Glc-β-amyrin)	m/z 627 (3/28-*O*-Glc-erythrodiol)	m/z 625 (3-*O*-Glc-OAld)	m/z 641 (3/28-*O*-Glc-OA)	m/z 773 (3-*O*-[Gal-Glc] -β-amyrin)	m/z 789 (3-*O*-[Gal-Glc] -erythrodiol)	m/z 787 (3-*O*-[Gal-Glc] -OAld)	m/z 803 (3-*O*-[Gal-Glc] -OA)
MD3	UGT73C10	-	-	-	-	-	-	-	-
MD5	UGT71G1	-	-	-	-	-	-	-	-
MD6	UGT73K1	-	-	-	-	-	-	-	-
MD10	UGT73C10 + UGT73P2	-	-	-	-	-	-	-	-
MD14	CYP716A12 + UGT73C10	13.4’	10.8’ (3) 10.9’ (28?)	11.8’	10.9’ (3)	-	9.9’ (Glc-Glc)	-	9.6’ (Glc-Glc)
MD15	CYP716A12 + UGT73F3	-	10.9’ (28?)	-	10.6’ (28)	-	9.9’ (Glc-Glc)	-	9.6’ (Glc-Glc)
MD16	CYP716A12 + UGT71G1	-	-	-	-	-	-	-	-
MD17	CYP716A12 + UGT73K1	-	-	-	-	-	-	-	-
MD21	CYP716A12 + UGT73C10 + UGT73P2	13.4’	10.8’ (3) 10.9’ (28?)	11.8’	10.9’ (3)	-	9.9’ (Glc-Glc) 10.5’ (Gal-Glc)	-	9.6’ (Glc-Glc) 10.5’ (Gal-Glc)

All strains expressed *EtBAS*. The m/z value refers to the sodium adduct of the parent ion, $[M+Na]^+$, of the corresponding saponin. Retention times (min) of the observed peaks are shown. The potential position of the sugar (3 or 28) or the hypothesised sugar chain (Glc-Glc, Gal-Glc) are shown in brackets. OAld: oleanolic aldehyde; OA: oleanolic acid; Glc: glucose; Gal: galactose; 3-*O*-[Gal-Glc]-OA: neosaponin

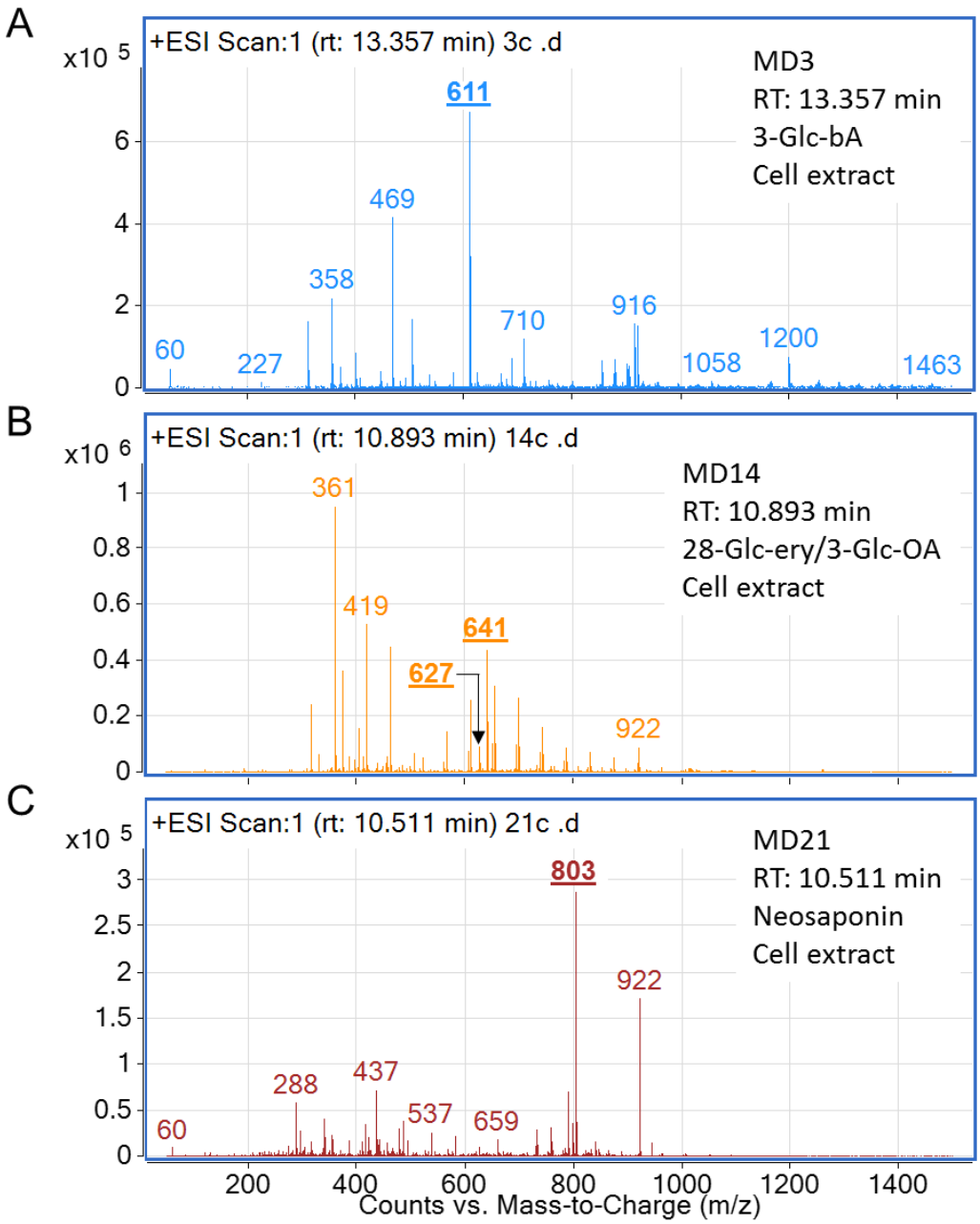


Figure 6.3. Mass spectra of selected peaks. Only the parent ions were visible, indicating that little or no fragmentation occurred. **(A)** Mass spectrum of the peak at 13.4 min in sample MD3 (cell extract), potentially corresponding to 3-*O*-Glc-β-amyrin. **(B)** Mass spectrum of the peak at 10.9 min of MD14 (cell extract), potentially corresponding to 28-*O*-Glc-erythrodiol (m/z 627 parent ion) and 3-*O*-Glc-oleanolic acid (m/z 641 parent ion). **(C)** Mass spectrum of the peak at 10.5 min in MD21 (cell extract), potentially corresponding to neosaponin (3-*O*-[Gal-Glc]-oleanolic acid). The putative saponin parent ions are in bold and underlined.

6.3.1 Analysis of saponin monoglycosides

Strains MD3 (expressing *EtBAS* and *UGT73C10*) and MD10 (*EtBAS*, *UGT73C10* and *UGT73P2*) had peaks potentially corresponding to 3-*O*-Glc-β-amyrin in the cell extract but not in the spent medium (Figure 6.4 A-B). By contrast, MD14 (oleanolic acid pathway plus *UGT73C10*) and MD21 (oleanolic acid pathway plus *UGT73C10* and *UGT73P2*) had such a peak in both the spent medium and the cell extract (Figure 6.4 A-B). There were small differences in the retention time of these peaks, which may result from experimental error. The other strains did not have any unique peaks, indicating that they did not produce 3-*O*-Glc-β-amyrin.

Analysis of EICs for the 627 m/z ion, which should correspond to glycosylated erythrodiol, showed two peaks (at 10.8 min and 10.9 min) for MD14 and MD21 and one peak (at 10.9 min) for MD15 (oleanolic acid pathway with *UGT73F3*) in both the cell extracts and spent media (Figure 6.4 C-D). While a small peak at 10.8 min was present in the cell extracts of all strains, this peak was considerably larger for MD14 and MD21 (Figure 6.4 C), indicating that it represented a metabolite unique to these strains. The peak at 10.8 min was the dominant peak in MD14 and MD21 and was not observed for MD15. UGT73C10 (expressed by MD14 and MD21) is reported to glycosylate the *C*-3 hydroxyl, while UGT73F3 (expressed by MD15) is reported to glycosylate only at *C*-28 and did not glycosylate β-amyrin (Figure 6.4 A-B), which has only a *C*-3 hydroxyl. This indicates that the peak at 10.8 min may correspond to 3-*O*-Glc-erythrodiol. The peak at 10.9 min might therefore result from the glycosylation of erythrodiol at the *C*-28 hydroxyl, with UGT73C10 showing some weak activity towards this reaction. Although UGT73C10 was only reported to have *C*-3 glycosylation activity, other members of the UGT73C subfamily from *Barbarea vulgaris* also performed *C*-28 glycosylation (Augustin et al., 2012; Erthmann et al., 2018). Both 3- and 28-*O*-Glc-erythrodiol would have the same molecular weight, and would therefore be observed in the same EIC.

A putative 3-*O*-Glc-oleanolic aldehyde peak (m/z 625) was observed in both the cell extracts and spent media for MD14 and MD21 at 11.8 min (Figure 6.4 E-F). Some additional peaks unique to these strains were also present in the spent media. In the cell extracts the peak at 11.8 min was more abundant for MD14 than MD21, while in the spent media it was more abundant for MD21. MD21 putatively produced saponin

diglycosides (see below), and it is possible that this enhanced secretion of saponins, resulting in this difference between MD14 and MD21.

The EICs for the m/z 641 ion, which would correspond to both 3- and 28-*O*-Glc-oleanolic acid, showed a peak at 10.9 min for MD14 and MD21 and at 10.6 min for MD15 (for both cell extract and spent media) (Figure 6.4 G-H). These may correspond to 3- and 28-*O*-Glc-oleanolic acid, respectively. If so, this would indicate that UGT73C10 glycosylated oleanolic acid only at the *C*-3 hydroxyl and not at *C*-28, while UGT73F3 only glycosylated the *C*-28 carboxyl. This contrasts with what was observed for glycosylated erythrodiol (Figure 6.4 C-D), where it appeared that UGT73C10 may have glycosylated erythrodiol at the *C*-28 hydroxyl. A peak at 10.9 min was also observed in the EICs of the m/z 627 ion for MD14, MD15 and MD21, which potentially corresponded to 28-*O*-Glc-erythrodiol (see above). In this case, 28-*O*-Glc-erythrodiol and 3-*O*-Glc-oleanolic acid would be co-eluting.

The results suggest that MD14 and MD21 potentially accumulate 3-*O*-Glc-β-amyrin, -erythrodiol, -oleanolic aldehyde and -oleanolic acid in both the spent media and within the cells, while MD3 and MD10 accumulate 3-*O*-Glc-β-amyrin only in the cells (Figure 6.4 A-H). This is consistent with the TLC results, where 3-*O*-Glc-β-amyrin was only detected in the spent medium in MD14 and MD21 (Figure 6.2), and suggests that the production of glycosylated erythrodiol, oleanolic aldehyde or oleanolic acid may allow the secretion of saponins into the growth media.

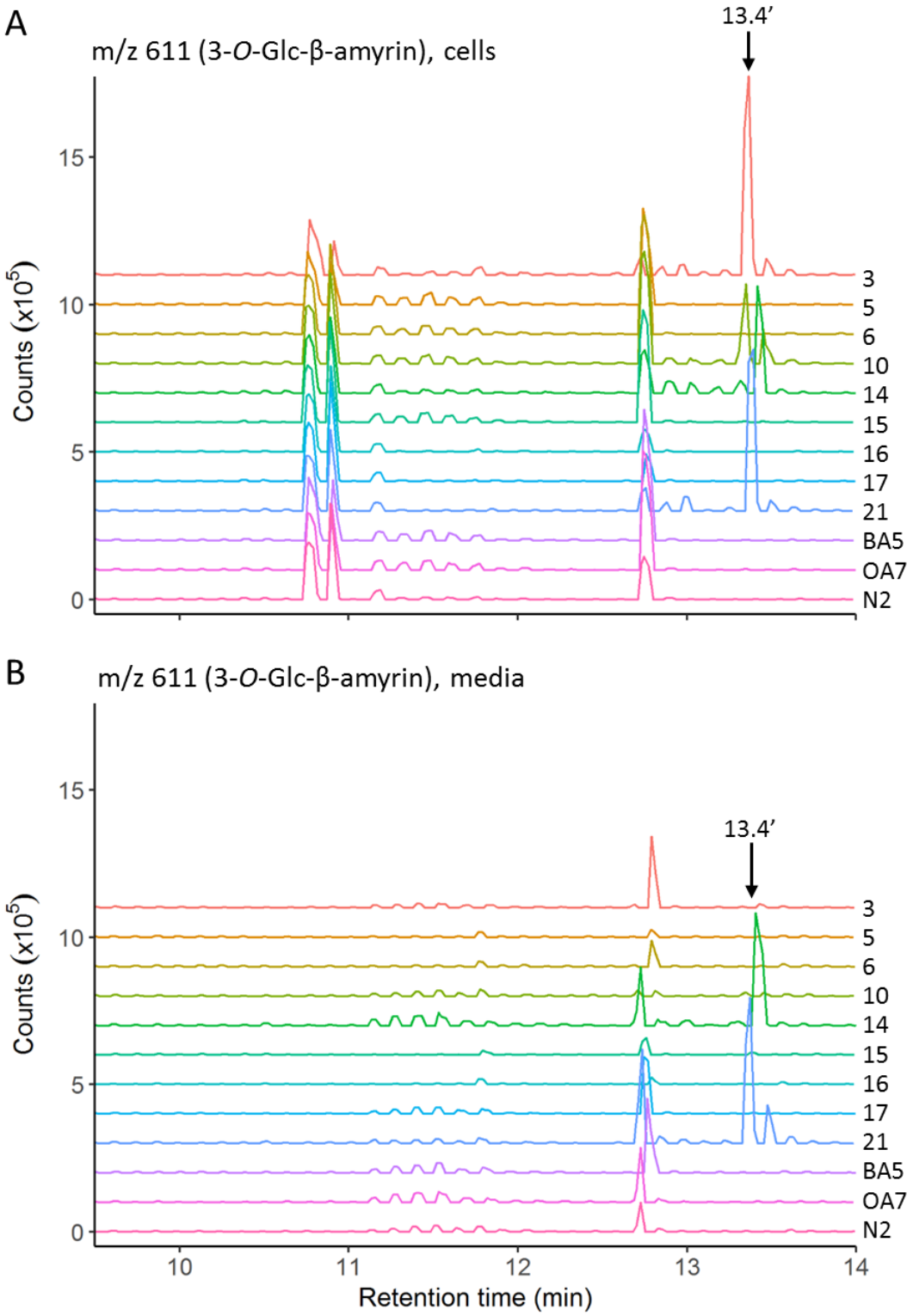
No saponins were detected for the strains expressing *UGT71G1* or *UGT73K1* in either the β-amyrin or oleanolic acid strains (MD5-6 and MD16-17). These UGTs were reported to glycosylate a variety of oxidised triterpenoids *in vitro*, but they were not tested against β-amyrin or its derivatives erythrodiol, oleanolic aldehyde or oleanolic acid (Achnine et al., 2005). The present results indicate that they do not have activity against these triterpenoids.

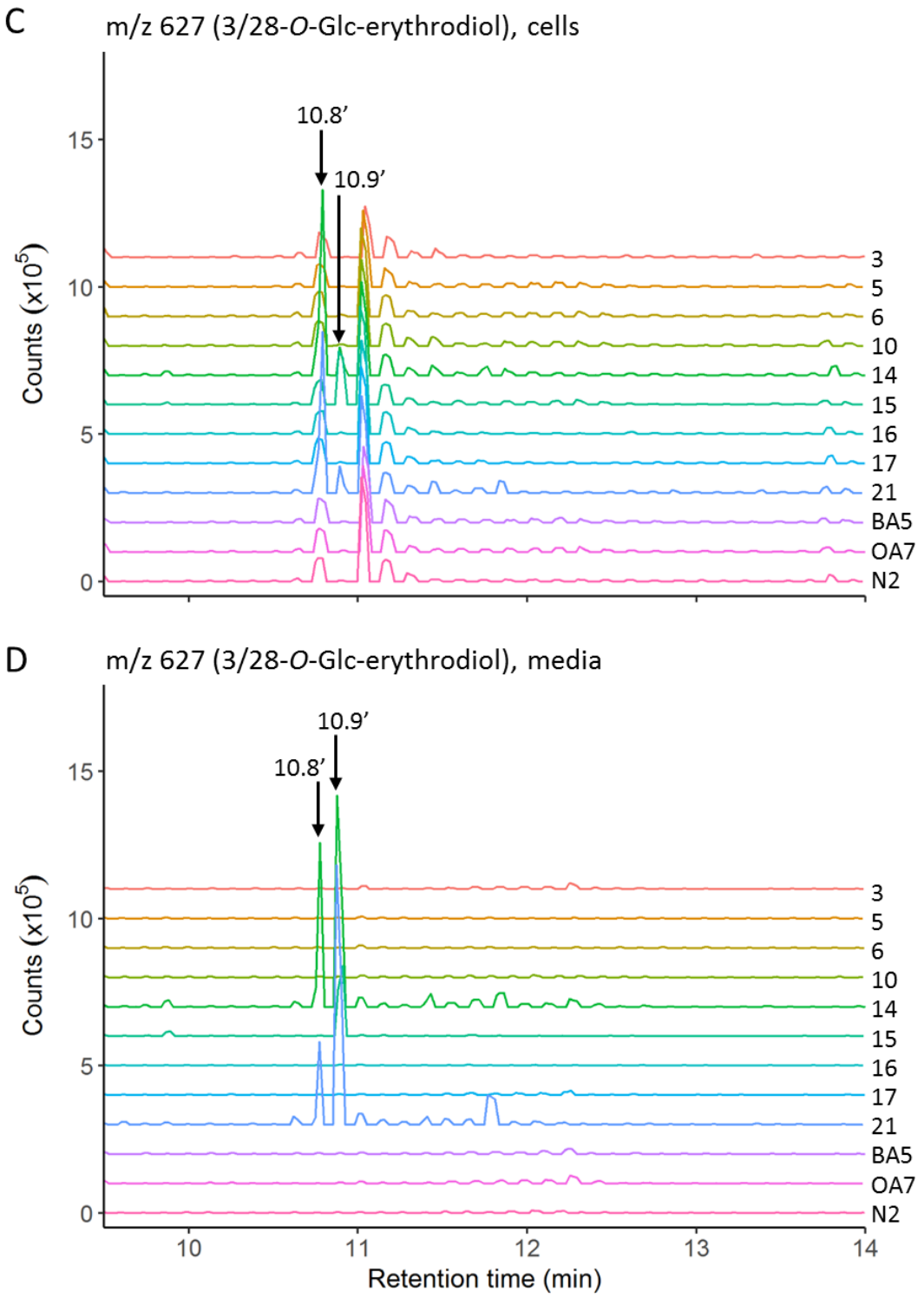
6.3.2 Analysis of saponin diglycosides

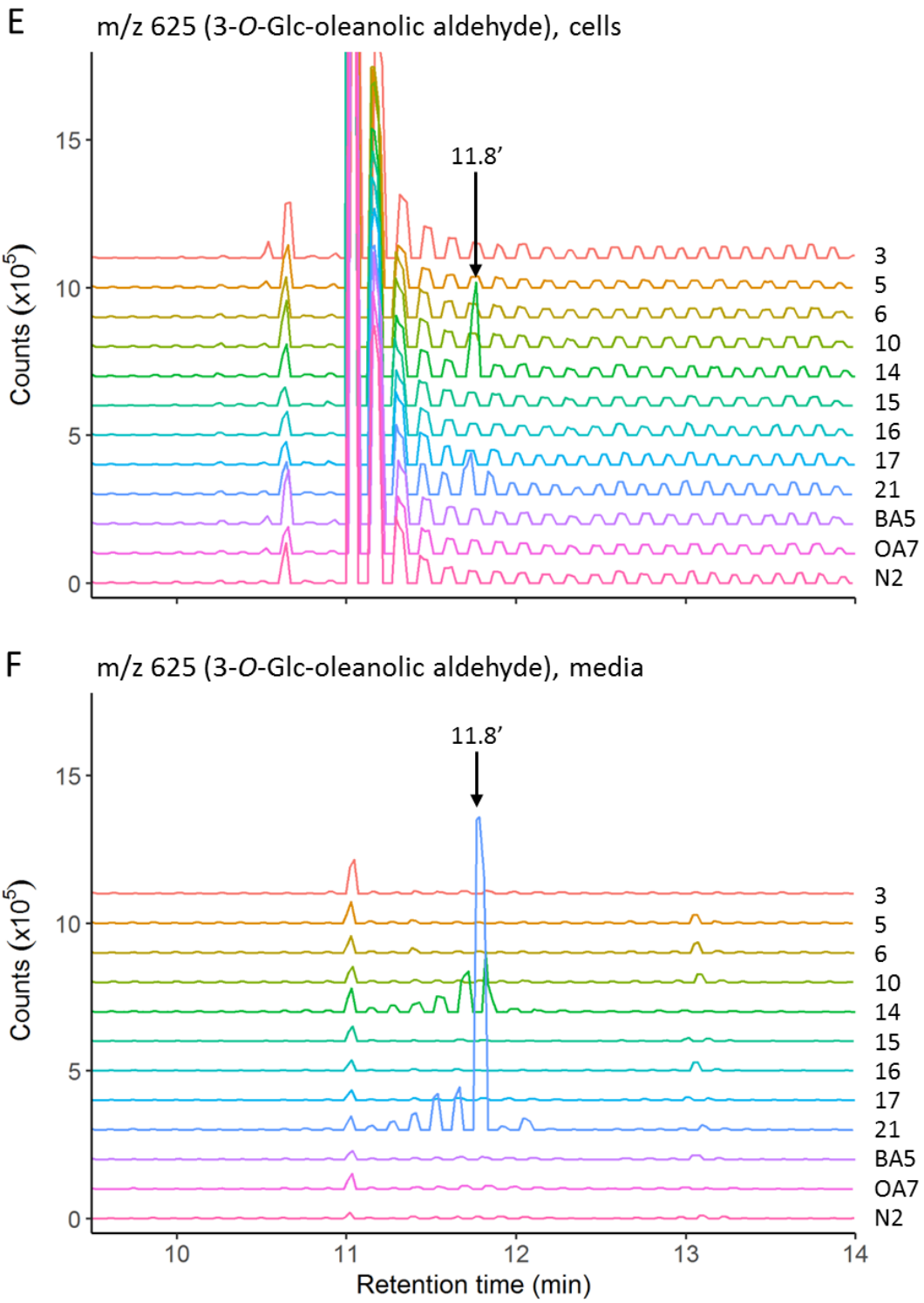
EICs corresponding to the saponin diglycosides 3-*O*-[Gal-Glc]-β-amyrin, -erythrodiol, -oleanolic aldehyde and -oleanolic acid (neosaponin) were also analysed. For MD21, two peaks (at 9.6 min and 10.5 min) were observed in the EIC for neosaponin in both the cell extracts and spent media (Figure 6.4 O-P). The first peak was in low abundance and was also observed for MD14 and MD15, suggesting that it may result

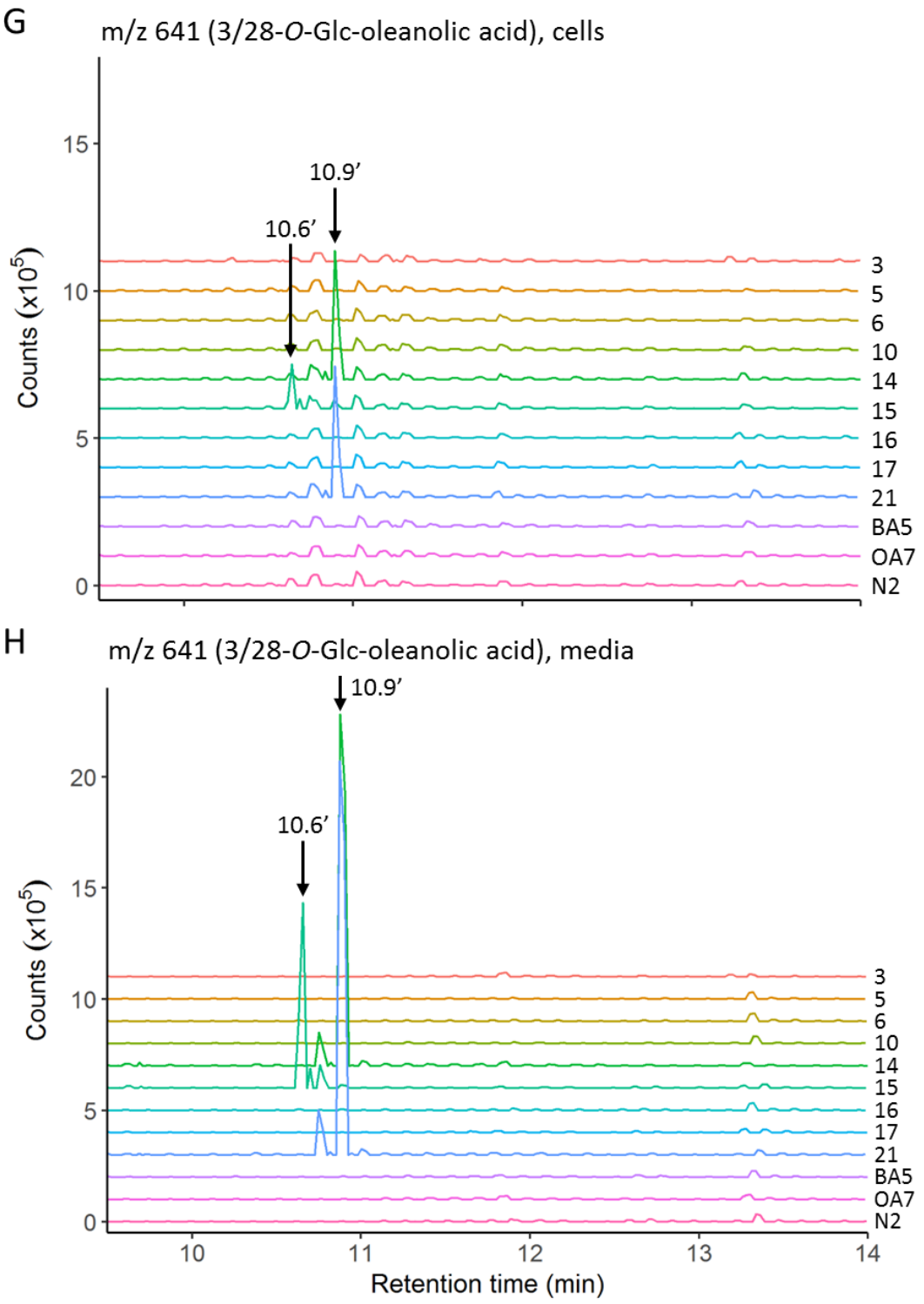
from the addition of two glucose molecules by UGT73C10 (for MD14 and MD21) or UGT73F3 (for MD15). Indeed, other members of the UGT73C from *Barbarea vulgaris* have been shown to add multiple glucose groups to their substrates (Augustin et al., 2012; Erthmann et al., 2018). Because glucose and galactose are isomers and therefore have the same molecular weight, oleanolic acid with a Glc-Glc sugar chain would give a parent ion with the same m/z value as a neosaponin, and both would therefore be observed in the same EIC.

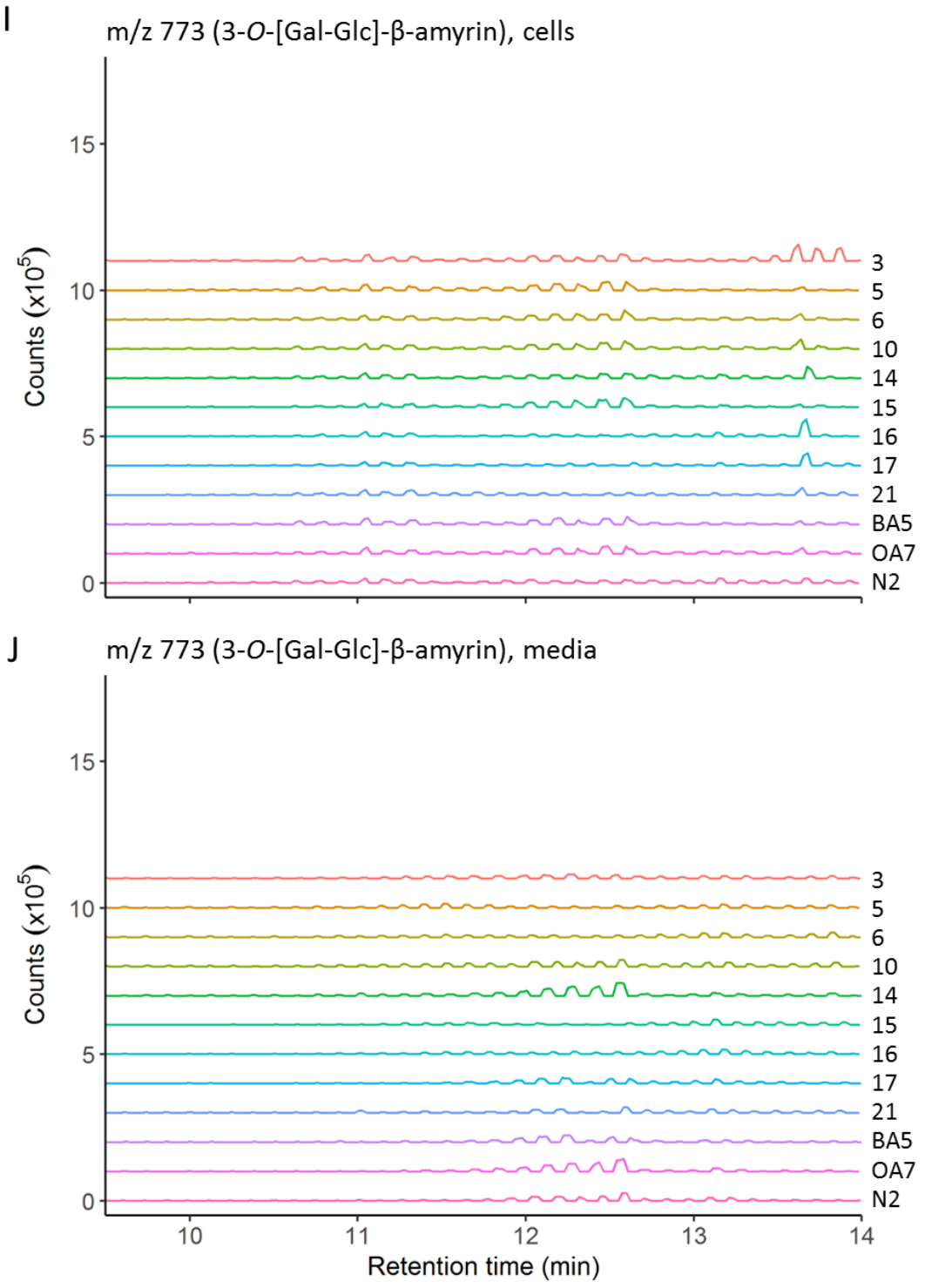
A similar result was observed with the EIC for 3-*O*-[Gal-Glc]-erythrodiol. MD21 had two peaks (at 9.9 min and 10.5 min) in both the cell extract and spent media, and the first peak was also observed for MD14 and MD15 (Figure 6.4 K-L). No peaks corresponding to 3-*O*-[Gal-Glc]-oleanolic aldehyde were observed for any strains (Figure 6.4 M-N). The EICs for this saponin showed many peaks that were also present in the control strains (MD-OA7 and the empty vector control MD-N2). This likely corresponds to noise from a background ion that has the same m/z as the 3-*O*-Glc-oleanolic aldehyde parent ion. No 3-*O*-[Gal-Glc]-β-amyrin was observed for MD10 or MD21. Because both MD10 and MD21 potentially produced 3-*O*-Glc-β-amyrin (Figure 6.4 A-B), this indicates that UGT73P2 may not be able to glycosylate this saponin.

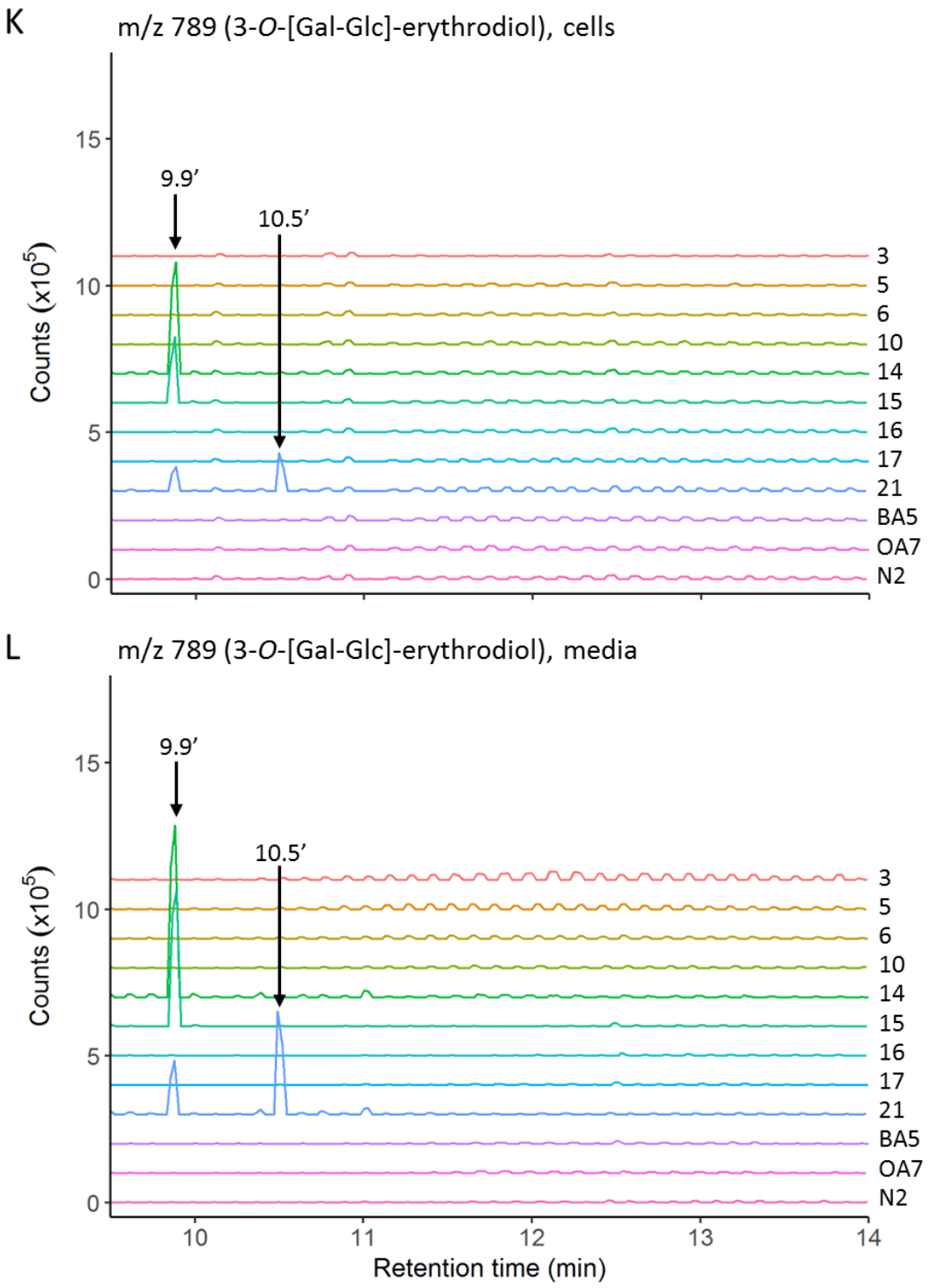

A
m/z 611 (3-O-Glc-β-amyrin), cells
13.4'
Counts (x10^5)
15
10
5
0
3
5
6
10
14
15
16
17
21
BA5
OA7
N2
B
m/z 611 (3-O-Glc-β-amyrin), media
13.4'
Counts (x10^5)
15
10
5
0
3
5
6
10
14
15
16
17
21
BA5
OA7
N2
10
11
12
13
14
Retention time (min)

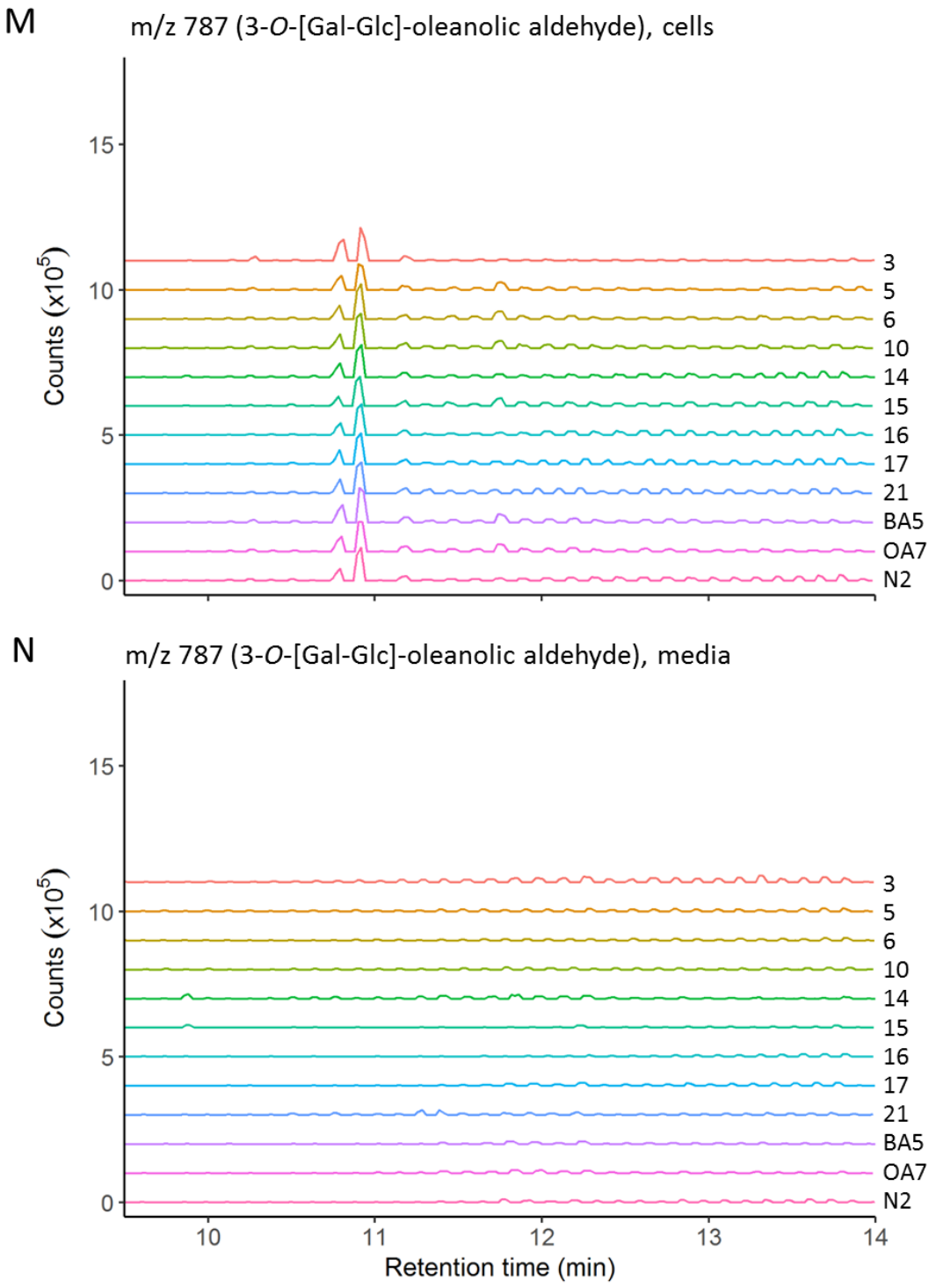

C
m/z 627 (3/28-O-Glc-erythrodiol), cells
10.8′
10.9′
Counts (x10^5)
15
10
5
0
3
5
6
10
14
15
16
17
21
BA5
OA7
N2
D
m/z 627 (3/28-O-Glc-erythrodiol), media
10.9′
10.8′
Counts (x10^5)
15
10
5
0
3
5
6
10
14
15
16
17
21
BA5
OA7
N2
10
11
12
13
14
Retention time (min)


E
m/z 625 (3-O-Glc-oleanolic aldehyde), cells
11.8'
Counts (x10^5)
15
10
5
0
3
5
6
10
14
15
16
17
21
BA5
OA7
N2
F
m/z 625 (3-O-Glc-oleanolic aldehyde), media
11.8'
Counts (x10^5)
15
10
5
0
3
5
6
10
14
15
16
17
21
BA5
OA7
N2
10
11
12
13
14
Retention time (min)


G
m/z 641 (3/28-O-Glc-oleanolic acid), cells
Counts (x10^5)
15
10
5
0
10.9'
10.6'
3
5
6
10
14
15
16
17
21
BA5
OA7
N2
H
m/z 641 (3/28-O-Glc-oleanolic acid), media
Counts (x10^5)
20
15
10
5
0
10.9'
10.6'
3
5
6
10
14
15
16
17
21
BA5
OA7
N2
10
11
12
13
14
Retention time (min)


I
m/z 773 (3-O-[Gal-Glc]-β-amyrin), cells
Counts (x10^5)
15
10
5
0
3
5
6
10
14
15
16
17
21
BA5
OA7
N2
J
m/z 773 (3-O-[Gal-Glc]-β-amyrin), media
Counts (x10^5)
15
10
5
0
3
5
6
10
14
15
16
17
21
BA5
OA7
N2
10
11
12
13
14
Retention time (min)


K
m/z 789 (3-O-[Gal-Glc]-erythrodiol), cells
9.9'
10.5'
Counts (x10^5)
15
10
5
0
3
5
6
10
14
15
16
17
21
BA5
OA7
N2
L
m/z 789 (3-O-[Gal-Glc]-erythrodiol), media
9.9'
10.5'
Counts (x10^5)
15
10
5
0
3
5
6
10
14
15
16
17
21
BA5
OA7
N2
10
11
12
13
14
Retention time (min)


M
m/z 787 (3-O-[Gal-Glc]-oleanolic aldehyde), cells
Counts (x10^5)
15
10
5
0
3
5
6
10
14
15
16
17
21
BA5
OA7
N2
N
m/z 787 (3-O-[Gal-Glc]-oleanolic aldehyde), media
Counts (x10^5)
15
10
5
0
3
5
6
10
14
15
16
17
21
BA5
OA7
N2
10
11
12
13
14
Retention time (min)

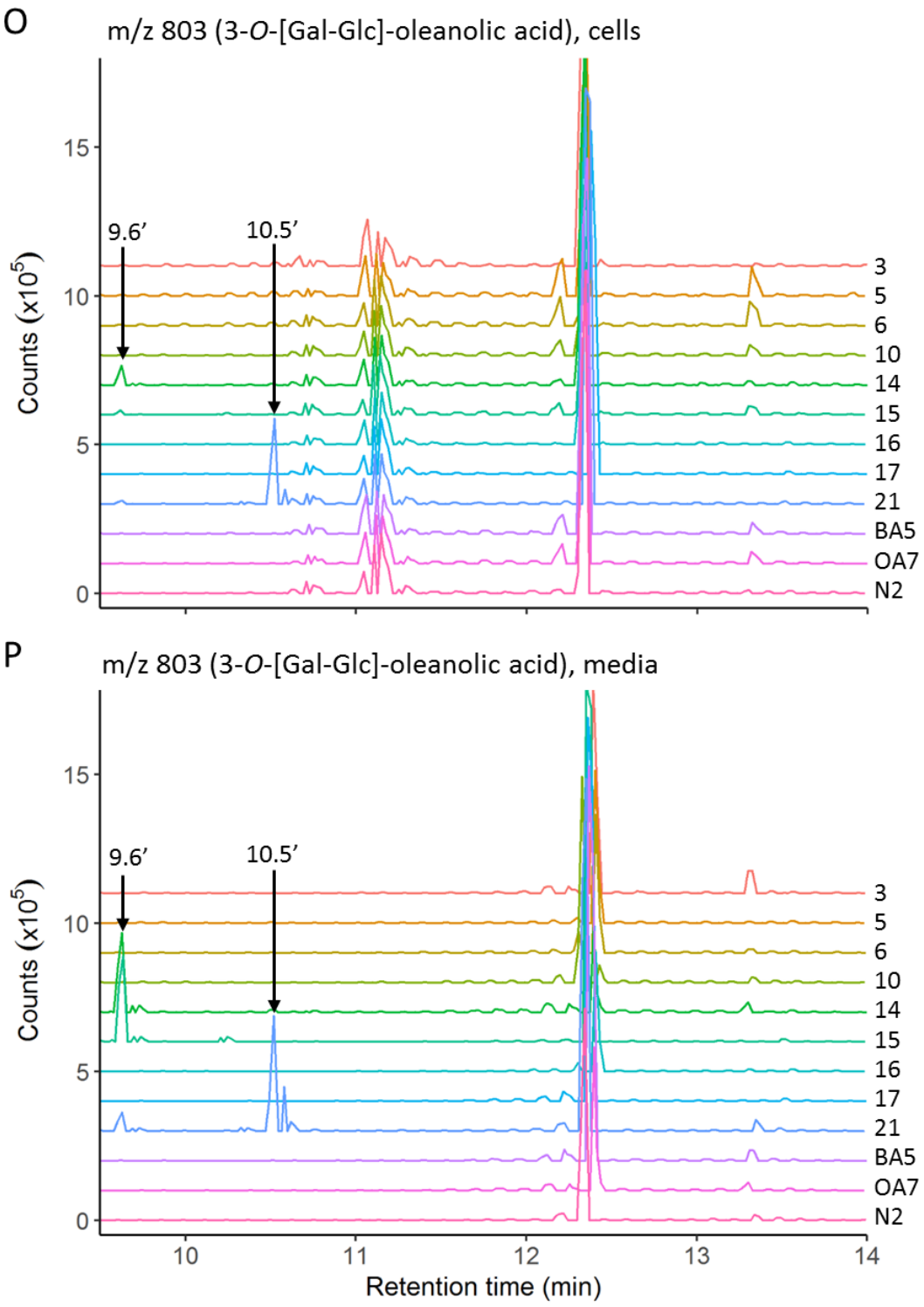


Figure 6.4. LC-MS analysis of strains engineered for saponin production. See next page for full figure legend.

Figure 6.4. LC-MS analysis of strains engineered for saponin production. Shown are overlays of the following extracted ion chromatograms (EICs) corresponding to Na^+ adducts: m/z 611, corresponding to 3-*O*-Glc-β-amyrin **(A-B)**; m/z 627, corresponding to 3- or 28-*O*-Glc-erythrodiol **(C-D)**; m/z 625, corresponding to 3-*O*-Glc-oleanolic aldehyde **(E-F)**; m/z 641, corresponding to 3- or 28-*O*-Glc-oleanolic acid **(G-H)**; m/z 773, corresponding to 3-*O*-[Gal-Glc]-β-amyrin **(I-J)**; m/z 789, corresponding to 3-*O*-[Gal-Glc]-erythrodiol **(K-L)**; m/z 787, corresponding to 3-*O*-[Gal-Glc]-oleanolic aldehyde **(M-N)**; and m/z 803, corresponding to 3-*O*-[Gal-Glc]-oleanolic acid (i.e. neosaponin) **(O-P)**. Both extracts from the cells and the spent media are shown. Peaks potentially corresponding to saponins are labelled with their retention times (min). MD-BA5 expressed *EtBAS*, MD-OA7 expressed *EtBAS*, *CYP716A12* and *ATR2*, and MD-N2 was an empty vector control. See Table 6.2 and Table 6.3 for the enzymes expressed by the saponin production strains.

6.4 Discussion

Here, yeast strains were engineered to produce saponins derived from β-amyrin and oleanolic acid. TLC indicated that saponins were produced, and LC-MS analysis supported this conclusion. Importantly, it should be noted that the LC-MS analysis was not able to absolutely confirm the identities of the saponins, and therefore does not represent a final confirmation of saponin production. The saponins underwent little apparent fragmentation, meaning that they could only be identified on the basis of their parent ions and the presence of peaks that were absent in the negative controls. Furthermore, as this was a preliminary experiment, strains were run without replicates. The data also contained a large amount of noise. While such noise was observed in some of the EICs (e.g. see Figure 6.4 E), it was especially apparent when analysing the mass spectra, where many peaks that did not appear to represent fragment ions were present (Figure 6.3). As such, to confirm production of these saponins, the LC-MS method must be improved to reduce the noise and allow fragmentation to facilitate proper identification.

To further facilitate saponin identification, tandem mass spectrometry (LC-MS/MS) could be performed, which allows selected ions to be targeted for further fragmentation. These additional MS/MS spectra thus provide more information about the analyte, facilitating identification. Alkaline hydrolysis of saponin extracts should also allow glycosylation at *C*-3 and *C*-28 to be distinguished. This specifically cleaves ester bonds, resulting in the loss of the sugar chain at the *C*-28 carboxyl group.

(Augustin et al., 2012). As it does not hydrolyse ether bonds, the sugar chain at the *C*-3 hydroxyl is unaffected. Thus, depletion of saponins after alkaline hydrolysis would indicate that glycosylation had occurred at the *C*-28 carboxyl group. The use of standards corresponding to the expected saponins (e.g. of 3-*O*-Glc-oleanolic acid etc.) would allow absolute identification of these saponins by LC-MS. Unfortunately, such standards were not available for the present study. In the future, these could be produced through the glycosylation of commercially available triterpenoid aglycones as described by Augustin *et al.*, incubating the aglycone with extracts of *E. coli* expressing the relevant UGT (Augustin et al., 2012).

The ability to identify saponins through analysis of TICs allows untargeted analyses to be performed. Because the fragment ions must be specified when using EICs, they are generally limited to targeted approaches in which the expected metabolites are known. It is therefore difficult to use EICs (and thus the method used in the present study) to detect unexpected compounds because it requires a fragment ion for every possible compound to be specified, meaning that any unexpected compounds would be missed. The ability to perform untargeted analyses would thus greatly expand the range of saponins that can be detected. This would enable combinatorial experiments where different biosynthetic enzymes are 'mixed-and-matched' in novel combinations, potentially allowing the production of a diverse range of saponins.

The production of neosaponin reported here represents the first production of an oleanane diglycoside saponin in yeast, and is an important step towards the production of complex bioactive saponins in this microorganism. However, the saponins produced in this work were simple in structure, containing only one or two sugar moieties and being oxygenated only at up to a single position (the *C*-28 moiety). Many bioactive saponins are far more complex, containing multiple sugar chains of different lengths, composed of multiple sugars, and with varying branching patterns (Bang et al., 2005; Moses et al., 2014a; Augustin et al., 2011; Ragupathi et al., 2011). Furthermore, saponins are often oxygenated at multiple positions and are sometimes modified by other tailoring enzymes (e.g. methyl- or acyltransferases) (Moses et al., 2014a; Thimmappa et al., 2014). QS-21 provides a great example of such a complex saponin, containing a branched trisaccharide at the *C*-3 hydroxyl, an acylated linear tetrasaccharide at the *C*-28 carboxyl, being oxygenated at three positions on the aglycone, and containing a total of six different sugars (Ragupathi et al., 2011).

Pertinently, the co-expression of *UGT73C10* and *UGT73F3* may allow the production of bidesmosidic saponins glycosylated at both the *C*-3 and *C*-28 positions. Production of such saponins has not been demonstrated in yeast before. Six UGTs of the UGT73C subfamily were recently identified from *Barbarea vulgaris* (Erthmann et al., 2018). *In vitro* characterisation using *E. coli* extracts expressing the UGTs indicated that four were able to glycosylate oleanolic acid or hederagenin, the *C*-23 hydroxy derivative of oleanolic acid, at both the *C*-3 hydroxyl and *C*-28 carboxyl groups. These are therefore also excellent candidates for the production of bidesmosides in yeast.

The co-expression of three or more UGTs for the production of saponins with longer sugar chains should also be explored. For example, soyasaponin I, which contains a rhamnose-galactose-glucuronic acid sugar chain at the *C*-3 hydroxyl, might be produced through the co-expression of *UGT73P2*, *UGT91H4* and a glucuronosyltransferase (transferring a glucuronic acid moiety). UGT73P2 and UGT91H4 are involved in soyasaponin I biosynthesis in *Glycine max*, with UGT91H4 transferring a rhamnose to the existing Gal-GlcA sugar chain (Shibuya et al., 2010). Although a glucuronosyltransferase has yet to be identified, two human UGTs were found to transfer a glucuronic acid to the *C*-3 or *C*-28 moiety of the triterpenoid ursolic acid (regiospecificity undetermined) (Gao et al., 2016). UGT91H4 might also be able to glycosylate neosaponin, which has a Gal-Glc sugar chain.

Importantly, glycosylation of triterpenoids requires the appropriate UDP-sugar, which is used as a donor by the UGT that catalyses the reaction (Thimmappa et al., 2014; Augustin et al., 2011). While saponins can contain a variety of different sugars, *S. cerevisiae* natively produces only three UDP-sugars: UDP-glucose, UDP-galactose and UDP-*N*-acetylglucosamine (Oka and Jigami, 2006). Indeed, QS-21 contains the sugars xylose, glucuronic acid, galactose, fucose, rhamnose and apiose (Ragupathi et al., 2011), meaning that the pathways for five UDP-sugars would need to be expressed in addition to the saponin biosynthetic pathway. It is therefore important to develop yeast strains capable of producing such UDP-sugars. As plants are known to encode a huge number of UGTs (Lairson et al., 2008; Caputi et al., 2012), they are likely an excellent source of UDP-sugar biosynthetic genes. Indeed, the pathways for UDP-glucuronic acid and UDP-xylose have successfully been expressed in yeast, using genes from *Arabidopsis thaliana* (Oka and Jigami, 2006). Future work should

thus explore the co-expression of UDP-sugar and saponin biosynthetic pathways to produce saponins containing a variety of different sugars.

Chapter 7: Final discussion

The aim of this work was to produce triterpenoids in *Saccharomyces cerevisiae*. Yeast natively produces the triterpene precursor 2,3-oxidosqualene but to produce triterpenoids a heterologous pathway must be expressed. Triterpenoid biosynthesis comprises three key steps: (i) cyclisation of 2,3-oxidosqualene by an OSC, producing a triterpene; (ii) oxygenation of the triterpene by cytochromes P450 to produce a triterpenoid; and (iii) glycosylation of the triterpenoid by UGTs, producing a triterpenoid saponin (Thimmappa et al., 2014). Additional tailoring reactions (e.g. acylations) can also occur (Thimmappa et al., 2014) but these were not the focus of this research.

In Chapter 3, a GC-MS method to monitor and quantify the production of triterpene and triterpenoid aglycones was presented. In Chapter 4, twelve β-amyrin synthases (BASs) were compared for their ability to cyclise 2,3-oxidosqualene to β-amyrin when expressed in yeast, and the effect on growth rate was assessed. Similarly, in Chapter 5, sixteen P450s of the CYP716A subfamily were compared for their ability to oxidise β-amyrin to oleanolic acid, and different combinations of BAS and CYP716A were also analysed. Finally, in Chapter 6, UGTs were additionally expressed in yeast engineered to produce β-amyrin or oleanolic acid, and saponin production was monitored by TLC and LC-MS.

In this chapter the following are discussed: (i) the impact of the BAS and CYP716A homologue on growth and productivity, (ii) understanding the differences in activities and the causes of burden, and (iii) the production of oleanane saponins in yeast.

7.1 The choice of homologue is important for strain development

Chapters 4-5 highlighted the benefits of comparing enzyme variants during strain development. The comparison of BASs found that β-amyrin titres ranged over 10-fold, while in the comparative analysis of CYP716As erythrodiol and oleanolic acid titres ranged 18.7- and 6.8-fold. Product ratios also varied depending on the homologues used. For example, strain MD-OA15 expressing *CYP716AL1* from *Catharanthus roseus* accumulated the greatest amounts of oleanolic acid (14.1 mg/L) and oleanolic aldehyde (titre undetermined) (Figure 5.7 A-B). Meanwhile, the strain expressing *CYP716A75* produced a surprisingly high amount of erythrodiol (20.9 mg/L) and very

little oleanolic aldehyde or oleanolic acid (titres undetermined) (Figure 5.7 A-B), indicating that it is highly specific for *C*-28 hydroxylation when using β-amyrin as a substrate. These differences in activity were not predictable from analysis of the amino acid sequences.

Strains expressing combinations of BASs and CYP716As with high (*AaBAS*, *CYP716AL1*) and low (*SlBAS*, *CYP716A52v2*) productivities were compared (Figure 5.12). Following 96 h growth in SC-URA + 2 % D-galactose media, strain MD-OA15 expressing the most productive homologues (*AaBAS* and *CYP716AL1*, with *ATR2*) yielded 8.2 mg/L β-amyrin, 7.3 mg/L erythrodiol and 8.5 mg/L oleanolic acid. Comparing MD-OA15 with strain MD-OA30 (expressing *SlBAS* and *CYP716AL1* with *ATR2*), differing from MD-OA15 only in the BAS used, showed that switching to the low productivity SlBAS substantially reduced triterpenoid production; this strain yielded 2.5 mg/L β-amyrin, 1.7 mg/L erythrodiol and 2.9 mg/L oleanolic acid. Similarly, using a low productivity CYP716A also limited production, as both strains expressing *CYP716A52v2* (MD-OA16 and MD-OA31) produced low quantities of oleanolic aldehyde and oleanolic acid compared with *CYP716AL1*, regardless of the BAS expressed.

Strikingly, replacing *AaBAS* with the less burdensome *AsBAS* (strain MD-OA29) resulted in almost complete conversion of β-amyrin to oleanolic acid (Figure 5.14 A-B). Furthermore, this strain grew at a similar rate as the strain expressing *AsBAS* alone (and thus also at a similar rate to the empty vector control strain) (Figure 5.15 A). Thus, the data demonstrate that the BAS and CYP716A variant can have a major impact on strain performance.

Comparative studies of other enzymes have previously been reported. Moses *et al.* (2014b) compared eight P450s of the CYP93E subfamily, which oxidise β-amyrin at the *C*-24 methyl group. Similar to the present work, the P450s displayed a range of activities and product ratios. Although 24-hydroxy-β-amyrin was the major product in all cases, smaller amounts of 24-carboxy- and 24-oxo-β-amyrin were also produced. By contrast, the CYP716As tested here showed a greater variety of product ratios, with erythrodiol often being the major product instead of oleanolic acid. More recently, Zhu *et al.* (2018) compared two *C*-11 oxidase (*CYP88D6*, *Uni25647*) and two *C*-30 oxidase (*CYP72A154*, *CYP72A63*) P450s for the production of glycyrrhetinic acid, which is derived from β-amyrin through two oxygenations. The combination of

Uni25647 and *CYP72A63* increased production of the intermediate 11-oxo-β-amyrin but did not significantly increased glycyrrhetinic acid production (Zhu et al., 2018).

All oleanolic acid strains in the present study expressed the *Arabidopsis thaliana* CPR *ATR2*. *A. thaliana* CPRs are often used for functional expression of P450s in yeast (Urban et al., 1997; Zhao et al., 2016; Dai et al., 2013, 2014; Moses et al., 2014a; Arendt et al., 2017; Gavira et al., 2013; Wang et al., 2015), however, numerous other CPRs have been characterised. For example, CPRs from *Lotus japonicus*, *Medicago truncatula*, *Catharanthus roseus* and *Artemisia annua* have been used to supply electrons to plant P450s expressed in yeast (Seki et al., 2008; Miettinen et al., 2017; Brown et al., 2015; Ro et al., 2006). The interaction between the P450 and CPR and the electron transfer to the P450 (coupling) is important for catalytic activity (Zhao et al., 2016; Renault et al., 2014), and it is possible that CPRs other than ATR2 might confer higher activity. The optimal CPR might also depend on the P450s expressed, with some P450/CPR pairs having better coupling than others. Indeed, Yan *et al.* (2014) compared ATR2 and two CPRs from *Panax ginseng* (PgCPR1 and PgCPR2) for the production of the triterpenoid protopanaxadiol, which is derived from the triterpene dammarenediol-II by a P450-catalysed oxygenation. PgCPR1 and PgCPR2 showed similar activities as judged by the protopanaxadiol titre, with PgCPR1 being slightly greater, while ATR2 showed considerably lower activity (approximately 2-fold).

More recently, Zhu *et al.* (2018) compared six CPRs from *A. thaliana* (ATR1, ATR2), *M. truncatula* (MTR2, MTR3), *L. japonicus* (LjCPR1) and *Glycyrrhiza uralensis* (GuCPR1) for the production of glycyrrhetinic acid in yeast. The choice of CPR had a large impact on productivity, with MTR3 producing very little glycyrrhetinic acid and GuCPR1 increasing production 5.6-fold compared with the commonly used ATR1 (Zhu et al., 2018). Meanwhile, ATR2, which was used in the present study, gave the second highest glycyrrhetinic acid titre (Zhu et al., 2018). Furthermore, the strain expressing *GuCPR1* accumulated far less β-amyrin compared with *ATR1*, indicating that flux was increased through the P450-catalysed parts of the pathway (Zhu et al., 2018). These differences in productivities could be explained by differences in the interaction between P450 and CPR, and it is therefore possible that GuCPR1 would not be the optimal CPR for all P450s.

Zhao *et al.* (2018) compared four CPRs (ATR1, MtCPR, LjCPR and GuCPR) in a strain expressing *GgBAS* and *CYP716A12* for the production of oleanolic acid. They found that MtCPR gave the highest oleanolic acid titres. It is possible that using a CPR from the same species as the P450 can give higher production, as both MtCPR and CYP716A12 are from *M. truncatula.* Indeed, in the comparison of CPRs for glycyrrhetinic acid production performed by Zhu *et al.* (2018), the optimal CPR (GuCPR) was from the same species as the expressed P450s (*G. uralensis*). Furthermore, the comparison of ATR2, PgCPR1 and PgCPR2 by Yan *et al.* (2014), which found PgCPR1 and PgCPR2 (from *P. ginseng*) to have considerably higher activities than ATR2 (from *A. thaliana*), was performed with the *P. ginseng* P450 CYP716U1. These studies, together with the present work, highlight how comparing enzyme variants can be a useful strategy to optimise strains for triterpenoid production.

With the exception of MD-OA29, all oleanolic acid strains accumulated a significant amount of β-amyrin, indicating that the P450-catalysed reactions were rate-limiting. As suggested by the work described above, testing various CPRs could prove an effective strategy to relieve this bottleneck (Yan et al., 2014; Zhao et al., 2018; Zhu et al., 2018). Another possible strategy is to fuse the P450 and CPR. This may improve coupling, enhancing electron transfer from the CPR to the P450 and reducing the generation of reactive oxygen species through electron leakage (Zangar et al., 2004; Renault et al., 2014). Zhao *et al.* (2016) successfully applied this strategy to boost the production of protopanaxadiol. Fusion of CYP716U1 (C-12 hydroxylase) with ATR1 resulted in almost complete conversion of dammarenediol-II to protopanaxadiol, while the non-fused variants accumulated a significant amount of dammarenediol-II (Zhao et al., 2016).

Interestingly, the co-expression of *AsBAS* (instead of *AaBAS*) with *CYP716AL1* and *ATR2* (strain MD-OA29) resulted in almost complete conversion of β-amyrin to oleanolic acid (Figure 5.14 A-B). Expression of *AsBAS* alone had little effect on growth rate compared with an empty vector control strain (MD-N1), and MD-OA29 also grew very similarly to MD-N1 (Figure 5.15 A). This apparent lack of burden could explain the improved the conversion of β-amyrin to oleanolic acid. *AsBAS* may be expressed at a lower level than *AaBAS*, causing less burden on the ER (to which it may be localised) and so allowing for greater P450 activity (the P450s and *ATR2* are also

believed to be localised to the ER), increasing conversion of β-amyrin to oleanolic acid.

Expression of the CYP716As and *ATR2* had little effect on growth rate, which could be explained by low expression levels. Indeed, they were not observed by western blot (Figure 5.11 A) and preliminary proteomics experiments also indicated that *CYP716AL1* and *ATR2* might be expressed at lower levels than *AaBAS* (Table 5.3). However, to confirm this hypothesis, further work will be required to detect the CYP716A and ATR2 proteins and compare their levels with those of the BASs.

7.2 Understanding differences in activities and the causes of burden

As noted above, expression of the BAS variants resulted in a variety of growth rates. Most grew more slowly than the empty vector control strain (MD-N1), indicating that BAS expression caused considerable metabolic burden. Interestingly, AsBAS grew at a similar rate to MD-N1, and PgBAS1 and BvBAS also grew relatively quickly. There are various explanations for this burden, including due to diversion of resources towards BAS expression, burden on the unfolded protein response due to misfolding of the BAS, toxicity caused by accumulation of β-amyrin, or depletion of endogenous metabolites due to β-amyrin production (e.g. 2,3-oxidosqualene and sterols). Depletion of ergosterol is perhaps unlikely, because ergosterol levels were not noticeably depleted in any of the BAS strains at the end of the 96 h culture period. However, it is possible that ergosterol levels were lower at earlier time points, only reaching levels similar to the wild-type by 96 h.

The smaller effect on growth rate of *AsBAS*, *PgBAS1* and *BvBAS* expression could possibly be explained by lower expression levels and/or less misfolding, leading to less burden on the unfolded protein response. Furthermore, these strains had relatively low specific β-amyrin titres (Figure 4.8 D), indicating that cellular β-amyrin levels were low in these strains. If β-amyrin accumulation is toxic, this might explain why these strains grew more quickly. However, while synthetic formiate and acetate derivatives of β-amyrin were found to inhibit growth of the yeast *Candida albicans* (Johann et al., 2007), to date there are no reports on the toxicity of β-amyrin to *S. cerevisiae*. Notably, the strain expressing *GhBAS* also had a low specific β-amyrin titre yet grew slowly (Figure 4.8 D, Figure 4.10 A), suggesting that other factors likely contribute to the differences in growth rate.

To investigate the causes of toxicity, the growth of a strain expressing a catalytically inactive variant of *AaBAS* (*dAaBAS*) was monitored. Any burden observed in this strain would not be expected to be caused by the accumulation of β-amyrin. This strain grew more quickly compared with wild-type AaBAS, reaching stationary phase after ~ 60 h instead of ~ 80 h, but more slowly than MD-N1 (which reached stationary phase after ~ 30 h) (Figure 4.15 A). This indicates that both production of the BAS protein and conversion of 2,3-oxidosqualene to β-amyrin slows yeast growth. However, it should be noted that expression levels were not accounted for. Decreased expression of *dAaBAS* would be expected to cause less burden and so could explain the observed results. Thus, to conclusively interpret these results, the relative expression levels of *dAaBAS* and *AaBAS* must be determined.

The additional expression of a CYP716A and *ATR2* had little impact on growth. Generally, strains expressing oleanolic acid pathways with different BASs grew at a similar rate to the corresponding strains expressing only a BAS (Figure 5.10 A, Figure 5.13 A, Figure 5.15 A). Thus, MD-OA29 (expressing *AsBAS*, *CYP716AL1* and *ATR2*) grew at a similar rate to the strain expressing *AsBAS* alone (and thus to MD-N1) (Figure 5.15 A). By contrast, MD-OA15 (expressing *AaBAS*, *CYP716AL1* and *ATR2*) grew much more slowly, at a similar rate to the strain expressing *AaBAS* alone (Figure 5.15 A).

The small effect on growth of CYP716A and *ATR2* expression could be due to lower expression levels of these enzymes compared with the BASs. Western blots were able to detect five BASs (out of six tested) but could not detect the CYP716As or ATR2 (Figure 4.11 A, Figure 5.11 A). Furthermore, preliminary proteomics experiments indicated that *CYP716AL1* and *ATR2* might be expressed at a lower level than *AaBAS* (Table 5.3). The *GAL3* promoter (*pGAL3*), from which *ATR2* was expressed, was reported to be weaker than the *GAL10* and *GAL1* promoters (*pGAL1* and *pGAL10*) (Paddon et al., 2013), possibly explaining the potentially low protein level of ATR2. However, both *pGAL1* and *pGAL10* are reported to have similar strengths (Weinhandl et al., 2014), so it was unexpected that CYP716AL1 (expressed from *pGAL10*) should be in lower abundance than AaBAS (expressed from *pGAL1*).

Thus, the relatively high expression of the BASs could be responsible for the observed reductions in growth, while the additional expression of *CYP716AL1* and *ATR2* would have little effect due to their low protein levels. The minimal effect on growth rate of

AsBAS expression could also be explained by low protein levels, and is possibly supported by the western blot which could not detect AsBAS despite the five other tested BASs being observed (Figure 4.11 A).

7.3 Production of oleanane saponins in yeast

Expression of various UGTs in yeast strains engineered to produce β-amyrin or oleanolic acid resulted in the tentative production of oleanane mono- and diglycoside saponins, as determined by TLC and LC-MS. In these preliminary experiments, saponin identities could not be confirmed due to limitations in the LC-MS method: fragmentation of the saponins was not observed (Figure 6.3), meaning that they could only be identified on the basis of their parent ion. While this allowed for putative confirmation of saponin production, confirmation of their identities would require optimisation of the method such that fragmentation can be observed, and ideally the use of authentic standards, which were not available for the current study.

Diglycosides of erythrodiol and oleanolic acid were tentatively observed by LC-MS. If confirmed, this would represent the first production of oleanane diglycoside saponins in yeast. Only five other oleanane saponins have been produced in yeast and all were monoglucosides (Moses et al., 2014b; Arendt et al., 2017). By contrast, production of the diglycoside dammarane-type saponin ginsenoside Rg3 in yeast was recently reported, with a yield of 3.5 µmol/g dry cell weight (Wang et al., 2015).

Due to the lack of available standards, saponin titres were not reported here. To date, titres for oleanane saponins produced in yeast have not been reported, and in fact relatively few studies have produced oleanane saponins in yeast (Moses et al., 2014b; Arendt et al., 2017). By contrast, ginsenosides, saponins that are derived from the tetracyclic dammarane triterpenoids, have been produced more extensively in yeast (Wang et al., 2015; Wei et al., 2015; Yan et al., 2014; Zhuang et al., 2017b). Titres of 93 mg/L and 42 mg/L were reported for ginsenosides Rh1 and F1, respectively, which are derived from the triterpene dammarenediol-II through the action of two P450s and a UGT (Wei et al., 2015). Meanwhile, a titre of 292 mg/L was reported for ginsenoside Rh2, which is produced from dammarenediol-II through the action of one P450 and one UGT (Zhuang et al., 2017b).

7.4 Future work

7.4.1 Determining protein levels and investigating differences in activities

This work has highlighted the differences in activities between BAS and CYP716A homologues. However, more work is needed to determine the reasons for these differences. As noted above, expression levels could explain the observed differences in activities and growth rates of the strains, yet these could not be determined through the course of this work. Preliminary western blot and proteomics experiments indicated that AsBAS may be expressed at a lower level than the other BASs, and that the CYP716As and ATR2 may also be expressed at lower levels. Nevertheless, a more comprehensive analysis is required to investigate the difference in protein expression levels between the BAS and CYP716A homologues. This is especially important for interpreting the results of the catalytically inactive AaBAS (dAaBAS) experiment, which indicated that β-amyrin production may indeed be a cause of burden. Relative protein levels might be effectively compared using an untargeted proteomics approach, which may also allow the observation of low abundance proteins that cannot be detected by western blot. Specific triterpenoid production could then be determined by dividing triterpenoid titres by relative protein expression, as proposed by Hoshino, (2017).

Following determination of relative protein levels, the generation of homology and substrate docking models of homologues with different specific activities might allow these differences in activities to be investigated. Analysis of the BAS amino acid sequences identified several residues in AsBAS that could affect activity, and this analysis would be enhanced by knowledge of specific activity. By contrast, due to the low sequence conservation of P450s, no obvious amino acid variants could be identified for the CYP716As that might explain the differences in activities. Thus, for the CYP716As especially, homology and substrate docking models should be analysed. Any potentially important residues identified from these analyses could then be subjected to mutagenesis studies to confirm their importance for enzyme activity. Such studies might even identify mutants with enhanced catalytic activities or altered product ratios.

7.4.2 Metabolic and process engineering for enhanced triterpenoid production

As noted before, the strains used in the present study were not metabolically engineered to enhanced triterpenoid production, and expressed only the triterpenoid biosynthetic pathway. Triterpenoid titres have been boosted considerably using traditional metabolic engineering strategies, with oleanolic acid titres as high as 607 mg/L (in a fermenter) being reported (Zhao et al., 2018). This strain expressed *GgBAS*, *CYP716A12* and *MtCPR* (the oleanolic acid pathway), and further overexpressed the MEV and early sterol pathway genes *tHMG1*, *ERG9* and *ERG1*. The tHmg1p enzyme catalyses the reduction of HMG-CoA to mevalonate, the rate-limiting step of the mevalonate pathway, while *ERG9* and *ERG1* encode squalene synthase and squalene epoxidase, which are together responsible for the conversion of two FPP molecules to 2,3-oxidosqualene (see section 1.1.1). Their overexpression should thus boost flux towards 2,3-oxidosqualene, providing more substrate for triterpenoid biosynthesis. With the exception of *GgBAS* and *tHMG1*, these genes were expressed from galactose-inducible promoters (*pGAL1* or *pGAL10*), similar to the strains in the present work (Zhao et al., 2018). The *GAL1* and *GAL80* genes, which are responsible for metabolism of galactose and repression of the *GAL* promoters in the absence of galactose, were also knocked out, which should increase expression from the *GAL* promoters (Zhao et al., 2018). Titres of 186 mg/L oleanolic acid were obtained in shake flasks when grown in rich media supplemented with both galactose and glucose (with only 11.6 mg/L β-amyrin accumulating) (Zhao et al., 2018). Batch fermentation in a 5 L bioreactor then boosted titres to 607 mg/L oleanolic acid (Zhao et al., 2018).

In addition to overexpressing MEV and early sterol pathway genes, the repression of *ERG7* has also been shown to redirect flux towards triterpene biosynthesis (Kirby et al., 2008; Moses et al., 2014b). The *ERG7* gene encodes lanosterol synthase, which would compete with the triterpene synthase (e.g. β-amyrin synthase) for 2,3-oxidosqualene substrate. An alternative approach employed by Arendt *et al.* (2017) knocked out the *PAH1* gene (encoding phosphatidic acid phosphatase), which was previously shown to induce an expansion of the ER (Santos-Rosa et al., 2005). Because enzymes involved in the biosynthesis of triterpenoids are believed to localise to the ER, this should increase the capacity of yeast to express these enzymes and, indeed, levels of CYP716A12 protein were shown to be considerably higher in the

pah1 mutant strain (Arendt et al., 2017). This mutation increased β-amyrin titres 8-fold, triterpenoid titres up to 6-fold, and titres of the saponin 28-*O*-Glc-medicagenic acid 16-fold (Arendt et al., 2017). However, only titres of the triterpenoid medicagenic acid, whose biosynthesis involves three P450-catalysed oxygenations, were reported (27 mg/L) (Arendt et al., 2017).

Optimising the culture conditions is another promising strategy for increasing triterpenoid production. Moses *et al.* (2014a) explored the use of cyclodextrins to sequester triterpenoids into the culture media. Cyclodextrins are toroidal shaped molecules with hydrophobic interiors and hydrophilic exteriors that may allow them to encapsulate and solubilise the hydrophobic triterpenoids. Addition of cyclodextrins to the media should thus sequester the triterpenoids to the growth medium, preventing any toxicity or feedback inhibition that may arise from their intracellular accumulation. Indeed, the addition of methyl-β-cyclodextrin resulted in the accumulation of substantial amounts of triterpenoids in the culture medium, and increased β-amyrin titres ~ 1.5-fold (Moses et al., 2014b).

A conceptually similar strategy that has been applied to the production of the dammarane-type triterpenoid protopanaxadiol is two-phase fermentation (Dai et al., 2013; Zhao et al., 2016). Here, an organic solvent is added to the culture medium and acts to extract triterpenoids from the cells throughout the fermentation process. Using methyl oleate as the solvent, Dai *et al.* (2013) increased protopanaxadiol titres from 812 mg/L to 1189 mg/L. However, titres of dammarenediol-II, the triterpene precursor to protopanaxadiol, also increased from 776 mg/L to 1548 mg/L, indicating that extraction of this more hydrophobic triterpene was favoured. Zhao *et al.* (2016) observed a similar effect when using dodecane as the solvent, with protopanaxadiol titres actually decreasing from 1437 mg/L to 312 mg/L when two-phase fermentation was applied. Meanwhile, dammarenediol-II titres increased from 19.7 mg/L to 312 mg/L (Zhao et al., 2016). Thus, two-phase fermentation strategies should be used with caution, and careful attention should be paid to the choice of organic solvent to ensure that extraction of the product of interest is favoured.

Finally, using an extensive process engineering strategy, Czarnotta *et al.* (2017) were able to boost titres of betulinic acid, the *C*-28 carboxy derivative of the triterpene lupeol, from 28 mg/L to 182 mg/L. While growth in shake flasks with glucose as the carbon source produced 28 mg/L betulinic acid, the use of a fermenter and the

application of an ethanol feed doubled titres to 57 mg/L. Subsequently, the use of nitrogen-free resting cell fermentations increased titres to 182 mg/L. Here, cells were first grown to high biomass with glucose, before being transferred to nitrogen-free media with an ethanol feed. The cells were unable to grow in the nitrogen-free media, and the increase in betulinic acid titres might be explained by the resulting redirection of metabolic flux from biomass generation to triterpenoid production (Czarnotta et al., 2017).

Thus, numerous strategies have been successfully employed to boost the production of a variety of triterpenoids in yeast. These strategies should be applicable to enhance the production of oleanolic acid and its derived saponins in the strains presented in the current study.

7.4.3 Expanding saponin production in yeast

To further facilitate saponin identification, alkaline hydrolysis could be performed to distinguish between *C*-3 and *C*-28 glycosides. This results in the loss of the sugar chain at the *C*-28 carboxyl group, due to the specific cleavage of ester bonds (sugar chains at the *C*-3 hydroxyl are joined via an ether linkage, which is not cleaved) (Augustin et al., 2012). Absolute confirmation of saponin structure could be achieved by nuclear magnetic resonance spectroscopy (NMR) (Moses et al., 2014b).

Following confirmation of the oleanane saponins produced in the present study, there are numerous directions this work could be taken. Many saponins are far more complex than those produced here, containing multiple linear or branched sugar chains consisting of many different sugars (Thimmappa et al., 2014; Augustin et al., 2011). For example, QS-21 has two sugar chains containing six different sugars (glucuronic acid, xylose, galactose, fucose, rhamnose and apiose) (Ragupathi et al., 2011). By contrast, *S. cerevisiae* only natively produces three UDP-sugars: UDP-glucose, UDP-galactose and UDP-*N*-acetylglucosamine (Oka and Jigami, 2006). Indeed, only one of the sugars in QS-21 (UDP-galactose) is natively produced by yeast. Thus, to further expand saponin production in yeast, the pathways for UDP-sugar biosynthesis must be functionally expressed in yeast. Biosynthetic pathways for UDP-glucuronic acid and UDP-xylose from *Arabidopsis thaliana* have been functionally expressed in yeast, but these were not used for saponin production (Oka and Jigami, 2006). To date, only saponins containing glucose and galactose sugars

have been produced in yeast (Moses et al., 2014a; Arendt et al., 2017; Wei et al., 2015; Wang et al., 2015; Yan et al., 2014; Zhuang et al., 2017; present work).

7.5 Closing comments

In this work, a panel of twelve BASs were compared for the production of β-amyrin in yeast, in order to optimise this first committed step in the biosynthesis of oleanane triterpenoids and saponins. Subsequently, sixteen P450s of the CYP716A subfamily were compared for their ability to oxygenate β-amyrin to oleanolic acid, a key oleanane triterpenoid that is a precursor to numerous bioactive products. In both cases, activities varied widely, with AaBAS from *Artemisia annua* and CYP716AL1 from *Catharanthus roseus* producing the greatest amounts of β-amyrin and oleanolic acid, respectively. Furthermore, the co-expression of *AsBAS* from *Avena strigosa* and *CYP716AL1* resulted in almost complete conversion of β-amyrin to oleanolic acid, a result that may be linked to the negligible impact on growth caused by *AsBAS* expression. Finally, oleanane saponins derived from β-amyrin and its oxygenated derivatives erythrodiol, oleanolic aldehyde and oleanolic acid were produced in yeast. This work thus represents a step towards the commercial production of industrially and medically relevant oleanane triterpenoids and saponins. Ultimately, the ability to produce complex saponins such as QS-21 will require the co-expression of multiple P450s and UGTs, various UDP-sugar biosynthesis enzymes, and undoubtedly extensive metabolic engineering and process optimisation to enhance production to commercially relevant levels.

References

Achnine, L., Huhman, D. V., Farag, M.A., Sumner, L.W., Blount, J.W., and Dixon, R.A. (2005). Genomics-based selection and functional characterization of triterpene glycosyltransferases from the model legume Medicago truncatula. Plant J. **41**: 875–887.

Aiba, Y., Watanabe, T., Terasawa, Y., Nakano, C., and Hoshino, T. (2018). Strictly Conserved Residues in Euphorbia tirucalli b-Amyrin Cyclase:Trp612 Stabilizes Transient Cation through Cation–p Interaction and CH–p Interaction of Tyr736 with Leu734 Confers Robust Local Protein Architecture. ChemBioChem **19**: 486–495.

Apel, A.R., D'Espaux, L., Wehrs, M., Sachs, D., Li, R.A., Tong, G.J., Garber, M., Nnadi, O., Zhuang, W., Hillson, N.J., Keasling, J.D., and Mukhopadhyay, A. (2017). A Cas9-based toolkit to program gene expression in Saccharomyces cerevisiae. Nucleic Acids Res. **45**: 496–508.

Arendt, P., Miettinen, K., Pollier, J., De Rycke, R., Callewaert, N., and Goossens, A. (2017). An endoplasmic reticulum-engineered yeast platform for overproduction of triterpenoids. Metab. Eng. **40**: 165–175.

Augustin, J.M., Drok, S., Shinoda, T., Sanmiya, K., Nielsen, J.K., Khakimov, B., Olsen, C.E., Hansen, E.H., Kuzina, V., Ekstrom, C.T., Hauser, T., and Bak, S. (2012). UDP-Glycosyltransferases from the UGT73C Subfamily in Barbarea vulgaris Catalyze Sapogenin 3-O-Glucosylation in Saponin-Mediated Insect Resistance. Plant Physiol. **160**: 1881–1895.

Augustin, J.M., Kuzina, V., Andersen, S.B., and Bak, S. (2011). Molecular activities, biosynthesis and evolution of triterpenoid saponins. Phytochemistry **72**: 435–457.

Back, A., Rossignol, T., Krier, F., Nicaud, J.M., and Dhulster, P. (2016). High-throughput fermentation screening for the yeast Yarrowia lipolytica with real-time monitoring of biomass and lipid production. Microb. Cell Fact. **15**: 1–12.

Bak, S., Beisson, F., Bishop, G., Hamberger, B., Höfer, R., Paquette, S., and Werck-Reichhart, D. (2011). Cytochromes P450. Arabidopsis Book **9**: e0144.

Bang, S.-C., Lee, J.-H., Song, G.-Y., Kim, D.-H., Yoon, M.-Y., and Ahn, B.-Z. (2005). Antitumor activity of Pulsatilla koreana saponins and their structure-activity relationship. Chem. Pharm. Bull. **53**: 1451–1454.

Blazeck, J., Garg, R., Reed, B., and Alper, H.S. (2012). Controlling promoter strength and regulation in Saccharomyces cerevisiae using synthetic hybrid promoters. Biotechnol. Bioeng. **109**: 2884–2895.

Blount, B.A., Driessen, M.R.M., and Ellis, T. (2016). GC preps: Fast and easy extraction of stable yeast genomic DNA. Sci. Rep. **6**: 1–4.

Brown, S., Clastre, M., Courdavault, V., and O'Connor, S.E. (2015). De novo production of the plant-derived alkaloid strictosidine in yeast. Proc. Natl. Acad.

Sci. **112**: 3205–3210.

Budzikiewicz, H., Wilson, J.M., and Djerassi, C. (1963). Mass Spectrometry in Structural and Stereochemical Problems. XXXII. Pentacyclic Triterpenes. J. Am. Chem. Soc. **85**: 3688–3699.

Caputi, L., Malnoy, M., Goremykin, V., Nikiforova, S., and Martens, S. (2012). A genome-wide phylogenetic reconstruction of family 1 UDP-glycosyltransferases revealed the expansion of the family during the adaptation of plants to life on land. Plant J. **69**: 1030–1042.

Castillo, D.A., Kolesnikova, M.D., and Matsuda, S.P.T. (2013). An effective strategy for exploring unknown metabolic pathways by genome mining. J. Am. Chem. Soc. **135**: 5885–5894.

Chen, L., Zhou, L., Wang, Y., Yang, G., Huang, J., Tan, Z., Wang, Y., Zhou, G., Liao, J., and Ouyang, D. (2017). Food and sex-related impacts on the pharmacokinetics of a single-dose of ginsenoside compound K in healthy subjects. Front. Pharmacol. **8**: 1–13.

Cheng, S., Du, Y., Ma, B., and Tan, D. (2009). Total synthesis of a furostan saponin, timosaponin BII. Org. Biomol. Chem. **7**: 3112.

ClinicalTrials.gov, Identifier NCT02255435 (2014). RTA 408 Capsul. Patients With Friedreich's Ataxia - MOXIe.

Curran, K.A., Crook, N.C., Karim, A.S., Gupta, A., Wagman, A.M., and Alper, H.S. (2014). Design of synthetic yeast promoters via tuning of nucleosome architecture. Nat. Commun. **5**: 1–8.

Curran, K.A., Karim, A.S., Gupta, A., and Alper, H.S. (2013). Use of expression-enhancing terminators in Saccharomyces cerevisiae to increase mRNA half-life and improve gene expression control for metabolic engineering applications. Metab. Eng. **19**: 88–97.

Czarnotta, E., Dianat, M., Korf, M., Granica, F., Merz, J., Maury, J., Baallal Jacobsen, S.A., Förster, J., Ebert, B.E., and Blank, L.M. (2017). Fermentation and purification strategies for the production of betulinic acid and its lupane-type precursors in Saccharomyces cerevisiae. Biotechnol. Bioeng. **114**: 2528–2538.

Dai, Z., Liu, Y., Zhang, X., Shi, M., Wang, B., Wang, D., Huang, L., and Zhang, X. (2013). Metabolic engineering of Saccharomyces cerevisiae for production of ginsenosides. Metab. Eng. **20**: 146–156.

Dai, Z., Wang, B., Liu, Y., Shi, M., Wang, D., Zhang, X., Liu, T., Huang, L., and Zhang, X. (2014). Producing aglycons of ginsenosides in bakers' yeast. Sci. Rep. **4**: 3698.

Dejong, J.M., Liu, Y., Bollon, A.P., Long, R.M., Jennewein, S., Williams, D., and Croteau, R.B. (2005). Genetic Engineering of Taxol Biosynthetic Genes in Saccharomyces cerevisiae. **4**.

Deng, S., Yu, B., Lou, Y., and Hui, Y. (1999). First total synthesis of an

exceptionally potent antitumor saponin, OSW- 1. J. Org. Chem. **64**: 202–208.

Didierlaurent, A.M., Laupèze, B., Di Pasquale, A., Hergli, N., Collignon, C., and Garçon, N. (2017). Adjuvant system AS01: helping to overcome the challenges of modern vaccines. Expert Rev. Vaccines **16**: 55–63.

Donald, K. a, Hampton, R.Y., and Fritz, I.B. (1997). Effects of overproduction of the catalytic domain of 3-hydroxy-3-methylglutaryl coenzyme A reductase on squalene synthesis in Saccharomyces cerevisiae . Effects of Overproduction of the Catalytic Domain of 3-Hydroxy-3-Methylglutaryl Coenzyme A Reductase o. **63**: 3341–3344.

Dong, L. et al. (2018). Co-expression of squalene epoxidases with triterpene cyclases boosts production of triterpenoids in plants and yeast. Metab. Eng. **49**: 1–12.

Erthmann, P.Ø., Agerbirk, N., and Bak, S. (2018). A tandem array of UDP-glycosyltransferases from the UGT73C subfamily glycosylate sapogenins, forming a spectrum of mono- and bisdesmosidic saponins. Plant Mol. Biol. **97**: 37–55.

Falade, O.S., Otemuyiwa, I.O., Oladipo, A., Oyedapo, O.O., Akinpelu, B.A., and Adewusi, S.R.A. (2005). The chemical composition and membrane stability activity of some herbs used in local therapy for anemia. J. Ethnopharmacol. **102**: 15–22.

Fiallos-Jurado, J., Pollier, J., Moses, T., Arendt, P., Barriga-Medina, N., Morillo, E., Arahana, V., de Lourdes Torres, M., Goossens, A., and Leon-Reyes, A. (2016). Saponin determination, expression analysis and functional characterization of saponin biosynthetic genes in Chenopodium quinoa leaves. Plant Sci. **250**: 188–197.

Field, B., Fiston-Lavier, a.-S., Kemen, a., Geisler, K., Quesneville, H., and Osbourn, a. E. (2011). Formation of plant metabolic gene clusters within dynamic chromosomal regions. Proc. Natl. Acad. Sci. **108**: 16116–16121.

Field, B. and Osbourn, A.E. (2008). Metabolic Diversification—Independent Assembly of Operon-Like Gene Clusters in Different Plants. **320**: 543–547.

Field, R.B. and Holmlund, C.E. (1977). Isolation of 2,3;22,23-Dioxidosqualene and 24,25-Oxidolanosterol from Yeast. Arch. Biochem. Biophys. **180**: 465–471.

Fiore, C., Eisenhut, M., Krausse, R., Ragazzi, E., Pellati, D., Armanini, D., and Bielenberg, J. (2008). Antiviral Effects of Glycyrrhiza species. Phytotherapy **22**: 141–148.

Fukushima, E.O., Seki, H., Ohyama, K., Ono, E., Umemoto, N., Mizutani, M., Saito, K., and Muranaka, T. (2011). CYP716A subfamily members are multifunctional oxidases in triterpenoid biosynthesis. Plant Cell Physiol. **52**: 2050–2061.

Fukushima, E.O., Seki, H., Sawai, S., Suzuki, M., Ohyama, K., Saito, K., and Muranaka, T. (2013). Combinatorial biosynthesis of legume natural and rare

triterpenoids in engineered yeast. Plant Cell Physiol. **54**: 740–749.

Gao, R., Liu, M., Chen, Y., Xia, C., Zhang, H., Xiong, Y., and Huang, S. (2016). Identification and characterization of human UDP-glucuronosyltransferases responsible for the in vitro glucuronidation of ursolic acid. Drug Metab. Pharmacokinet. **31**: 261–268.

Gauthier, C., Legault, J., Girard-Lalancette, K., Mshvildadze, V., and Pichette, A. (2009). Haemolytic activity, cytotoxicity and membrane cell permeabilization of semi-synthetic and natural lupane- and oleanane-type saponins. Bioorganic Med. Chem. **17**: 2002–2008.

Gavira, C., Höfer, R., Lesot, A., Lambert, F., and Zucca, J. (2013). Challenges and pitfalls of P450-dependent (þ) -valencene bioconversion by Saccharomyces cerevisiae. Metab. Eng. **18**: 25–35.

Geisler, K., Hughes, R.K., Sainsbury, F., Lomonossoff, G.P., Rejzek, M., Fairhurst, S., Olsen, C.-E., Motawia, M.S., Melton, R.E., Hemmings, A.M., Bak, S., and Osbourn, A. (2013). Biochemical analysis of a multifunctional cytochrome P450 (CYP51) enzyme required for synthesis of antimicrobial triterpenes in plants. Proc. Natl. Acad. Sci. U. S. A. **110**: E3360-7.

Ghosh, S. (2016). Biosynthesis of structurally diverse triterpenes in plants: The role of oxidosqualene cyclases. Proc. Indian Natl. Sci. Acad. **82**: 1189–1210.

Ghosh, S. (2017). Triterpene Structural Diversification by Plant Cytochrome P450 Enzymes. Front. Plant Sci. **8**: 1–15.

Gosling, R. and von Seidlein, L. (2016). The Future of the RTS,S/AS01 Malaria Vaccine: An Alternative Development Plan. PLoS Med. **13**: 1–6.

Han, J.Y., Chun, J.H., Oh, S.A., Park, S.B., Hwang, H.S., Lee, H., and Choi, Y.E. (2018). Transcriptomic Analysis of Kalopanax septemlobus and Characterization of KsBAS, CYP716A94 and CYP72A397 Genes Involved in Hederagenin Saponin Biosynthesis. Plant Cell Physiol. **59**: 319–330.

Han, J.Y., Hwang, H.S., Choi, S.W., Kim, H.J., and Choi, Y.E. (2012). Cytochrome P450 CYP716A53v2 catalyzes the formation of protopanaxatriol from protopanaxadiol during ginsenoside biosynthesis in panax ginseng. Plant Cell Physiol. **53**: 1535–1545.

Haralampidis, K., Bryan, G., Qi, X., Papadopoulou, K., Bakht, S., Melton, R., and Osbourn, a (2001). A new class of oxidosqualene cyclases directs synthesis of antimicrobial phytoprotectants in monocots. Proc. Natl. Acad. Sci. U. S. A. **98**: 13431–13436.

Hayashi, H., Huang, P., Kirakosyan, A., Inoue, K., Hiraoka, N., Ikeshiro, Y., Kushiro, T., Shibuya, M., and Ebizuka, Y. (2001). Cloning and characterization of a cDNA encoding beta-amyrin synthase involved in glycyrrhizin and soyasaponin biosyntheses in licorice. Biol. Pharm. Bull. **24**: 912–916.

Horwitz, A.A., Walter, J.M., Schubert, M.G., Kung, S.H., Hawkins, K., Platt, D.M.,

Hernday, A.D., Mahatdejkul-Meadows, T., Szeto, W., Chandran, S.S., and Newman, J.D. (2015). Efficient Multiplexed Integration of Synergistic Alleles and Metabolic Pathways in Yeasts via CRISPR-Cas. Cell Syst. **1**: 88–96.

Hoshino, T. (2017). β-Amyrin biosynthesis: catalytic mechanism and substrate recognition. Org. Biomol. Chem. **15**: 2869–2891.

Hoshino, T., Nakagawa, K., Aiba, Y., Itoh, D., Nakada, C., and Masukawa, Y. (2017). Euphorbia tirucalli β-Amyrin Synthase: Critical Roles of Steric Sizes at Val483 and Met729 and the CH–π Interaction between Val483 and Trp534 for Catalytic Action. ChemBioChem **18**: 2145–2155.

Hu, Z., He, B., Ma, L., Sun, Y., Niu, Y., and Zeng, B. (2017). Recent Advances in Ergosterol Biosynthesis and Regulation Mechanisms in Saccharomyces cerevisiae. Indian J. Microbiol. **57**: 270–277.

Huang, L., Li, J., Ye, H., Li, C., Wang, H., Liu, B., and Zhang, Y. (2012). Molecular characterization of the pentacyclic triterpenoid biosynthetic pathway in Catharanthus roseus. Planta **236**: 1571–1581.

Huber, R., Ritter, D., Hering, T., Hillmer, A.K., Kensy, F., Müller, C., Wang, L., and Büchs, J. (2009). Robo-Lector - A novel platform for automated high-throughput cultivations in microtiter plates with high information content. Microb. Cell Fact. **8**: 1–15.

Immethun, C.M., Hoynes-O'Connor, A.G., Balassy, A., and Moon, T.S. (2013). Microbial production of isoprenoids enabled by synthetic biology. Front. Microbiol. **4**: 1–8.

Ito, R., Masukawa, Y., and Hoshino, T. (2013). Purification, kinetics, inhibitors and CD for recombinant β-amyrin synthase from Euphorbia tirucalli L and functional analysis of the DCTA motif, which is highly conserved among oxidosqualene cyclases. FEBS J. **280**: 1267–1280.

Ito, R., Masukawa, Y., Nakada, C., Amari, K., Nakano, C., and Hoshino, T. (2014). β-Amyrin synthase from Euphorbia tirucalli. Steric bulk, not the π-electrons of Phe, at position 474 has a key role in affording the correct folding of the substrate to complete the normal polycyclization cascade. Org. Biomol. Chem. **12**: 3836–3846.

Ito, R., Nakada, C., and Hoshino, T. (2017). β-Amyrin synthase from Euphorbia tirucalli L. functional analyses of the highly conserved aromatic residues Phe413, Tyr259 and Trp257 disclose the importance of the appropriate steric bulk, and cation–π and CH–π interactions for the efficient catalytic ac. Org. Biomol. Chem. **15**: 177–188.

Iturbe-Ormaetxe, I., Haralampidis, K., Papadopoulou, K., and Osbourn, A.E. (2003). Molecular cloning and characterization of triterpene synthases from Medicago truncatula and Lotus japonicus. Plant Mol. Biol. **51**: 731–743.

Jacquier, N., Mishra, S., Choudhary, V., and Schneiter, R. (2013). Expression of oleosin and perilipins in yeast promotes formation of lipid droplets from the endoplasmic reticulum. J. Cell Sci. **126**: 5198–5209.

Jensen, K. and Møller, B.L. (2010). Plant NADPH-cytochrome P450 oxidoreductases. Phytochemistry **71**: 132–141.

Jin, M.L., Lee, D.Y., Um, Y., Lee, J.H., Park, C.G., Jetter, R., and Kim, O.T. (2014). Isolation and characterization of an oxidosqualene cyclase gene encoding a β-amyrin synthase involved in Polygala tenuifolia Willd. saponin biosynthesis. Plant Cell Rep. **33**: 511–519.

Johann, S., Soldi, C., Lyon, J.P., Pizzolatti, M.G., and Resende, M.A. (2007). Antifungal activity of the amyrin derivatives and in vitro inhibition of Candida albicans adhesion to human epithelial cells. Lett. Appl. Microbiol. **45**: 148–153.

Kajikawa, M., Yamato, K.T., Fukuzawa, H., Sakai, Y., Uchida, H., and Ohyama, K. (2005). Cloning and characterization of a cDNA encoding β-amyrin synthase from petroleum plant Euphorbia tirucalli L. Phytochemistry **66**: 1759–1766.

Kempinski, C., Jiang, Z., Bell, S., and Chappell, J. (2015). Metabolic engineering of higher plants and algae for isoprenoid production.

Kenny, P.T.M. and Wetzel, J.M. (1995). Letter: Fragmentation studies of ergosterol. The formation of the fragment ion at m/z 337. Eur. Mass Spectrom **1**: 411–413.

Kensy, F., Zang, E., Faulhammer, C., Tan, R.K., and Büchs, J. (2009). Validation of a high-throughput fermentation system based on online monitoring of biomass and fluorescence in continuously shaken microtiter plates. Microb. Cell Fact. **8**: 1–17.

Khakimov, B., Kuzina, V., Erthmann, P., Fukushima, E.O., Augustin, J.M., Olsen, C.E., Scholtalbers, J., Volpin, H., Andersen, S.B., Hauser, T.P., Muranaka, T., and Bak, S. (2015). Identification and genome organization of saponin pathway genes from a wild crucifer, and their use for transient production of saponins in Nicotiana benthamiana. Plant J. **84**: 478–490.

Kim, M., Lim, E., and Jung, M. (2013). First total synthesis of natural pulsatilla saponin D via highly stereospecific glycosylation. Tetrahedron **69**: 5481–5486.

Kirby, J. and Keasling, J.D. (2009). Biosynthesis of plant isoprenoids: perspectives for microbial engineering. Annu. Rev. Plant Biol. **60**: 335–355.

Kirby, J., Romanini, D.W., Paradise, E.M., and Keasling, J.D. (2008). Engineering triterpene production in Saccharomyces cerevisiae - β-amyrin synthase from Artemisia annua. FEBS J. **275**: 1852–1859.

Klug, L. and Daum, G. (2014). Yeast lipid metabolism at a glance. **14**: 369–388.

Komlaga, G. and Cojean, S. (2015). The Antimalarial Potential of Three Ghanaian Medicinal Plants. Herb. Med. Open Access **1**: 1–7.

Kotopka, B.J., Li, Y., and Smolke, C.D. (2018). Synthetic biology strategies toward heterologous phytochemical production. Nat. Prod. Rep. **00**: 1–19.

Kumar, S., Stecher, G., and Tamura, K. (2016). MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. Mol. Biol. Evol. **33**: 1870–

1874.

Kushiro, T., Shibuya, M., and Ebizuka, Y. (1998). β-Amyrin synthase. Cloning of oxidosqualene cyclase that catalyzes the formation of the most popular triterpene among higher plants. Eur. J. Biochem. **256**: 238–244.

Kushiro, T., Shibuya, M., Masuda, K., and Ebizuka, Y. (2000). Mutational studies on triterpene synthases: Engineering lupeol synthase into β-amyrin synthase. J. Am. Chem. Soc. **122**: 6816–6824.

Kuzina, V., Ekstrøm, C.T., Andersen, S.B., Nielsen, J.K., Olsen, C.E., and Bak, S. (2009). Identification of defense compounds in Barbarea vulgaris against the herbivore Phyllotreta nemorum by an ecometabolomic approach. Plant Physiol. **151**: 1977–1990.

Lairson, L.L., Henrissat, B., Davies, G.J., and Withers, S.G. (2008). Glycosyltransferases: Structures, Functions, and Mechanisms. Annu. Rev. Biochem. **77**: 521–555.

Leber, R., Zenz, R., Schröttner, K., Fuchsbichler, S., Pühringer, B., and Turnowsky, F. (2001). A novel sequence element is involved in the transcriptional regulation of expression of the ERG1 (squalene epoxidase) gene in Saccharomyces cerevisiae. Eur. J. Biochem. **268**: 914–924.

Lee, M.E., DeLoache, W.C., Cervantes, B., and Dueber, J.E. (2015). A Highly Characterized Yeast Toolkit for Modular, Multipart Assembly. ACS Synth. Biol. **4**: 975–986.

Liang, Y.L., Zhao, S.J., Xu, L.X., and Zhang, X.Y. (2012). Heterologous expression of dammarenediol synthase gene in an engineered Saccharomyces cerevisiae. Lett. Appl. Microbiol. **55**: 323–329.

Lin, C., Wen, X., and Sun, H. (2016). Oleanolic acid derivatives for pharmaceutical use: a patent review. Expert Opin. Ther. Pat. **26**: 643–655.

Liu, Q., Chen, W., Jiao, Y., Hou, J., Wu, Q., Liu, Y., and Qi, X. (2014). Pulsatilla saponin A, an active molecule from Pulsatilla chinensis, induces cancer cell death and inhibits tumor growth in mouse xenograft models. J. Surg. Res. **188**: 387–395.

Liu, Y., Zhao, Z., Xue, Z., Wang, L., Cai, Y., Wang, P., Wei, T., Gong, J., Liu, Z., Li, J., Li, S., and Xiang, F. (2016). An Intronless β-amyrin Synthase Gene is More Efficient in Oleanolic Acid Accumulation than its Paralog in Gentiana straminea. Sci. Rep. **6**: 1–13.

M'Baya, B., Fegueur, M., Servouse, M., and Karst, F. (1989). Regulation of squalene synthetase and squalene epoxidase activities in saccharomyces cerevisiae. Lipids **24**: 1020–1023.

Ma, Y., Gao, Y., Zhao, X., Zhu, Y., Du, F., and Hu, J. (2018). A Natural Triterpene Saponin-based Pickering Emulsion. Chem. - A Eur. J.

Madsen, K.M. (2011). Linking Genotype and Phenotype of Saccharomyces cerevisiae Strains Reveals Metabolic Engineering Targets and Leads to

Triterpene Hyper-Producers. PLoS One **6**.

Mahmoudi, S. and Keshavarz, H. (2017). Efficacy of phase 3 trial of RTS,S/AS01 malaria vaccine: The need for an alternative development plan. Hum. Vaccin. Immunother. **13**: 2098–2101.

Meesapyodsuk, D., Balsevich, J., Reed, D.W., and Covello, P.S. (2007). Saponin biosynthesis in Saponaria vaccaria. cDNAs encoding β-amyrin synthase and a triterpene carboxylic acid glucosyltransferase. Plant Physiol. **143**: 959–969.

Miettinen, K. et al. (2017). The ancient CYP716 family is a major contributor to the diversification of eudicot triterpenoid biosynthesis. Nat. Commun. **8**.

Milla, P., Athenstaedt, K., Viola, F., Oliaro-Bosso, S., Kohlwein, S.D., Daum, G., and Balliano, G. (2002). Yeast oxidosqualene cyclase (Erg7p) is a major component of lipid particles. J. Biol. Chem. **277**: 2406–2412.

Miziorko, H.M. (2011). Enzymes of the mevalonate pathway of isoprenoid biosynthesis. Arch. Biochem. Biophys. **505**: 131–143.

Mizutani, K., Kuramoto, T., Tamura, Y., Ohtake, N., Doi, S., Nakaura, M., and Tanaka, O. (1994). Sweetness of Glycyrrhetic Acid 3-O-β-D-Monoglucuronide and the Related Glycosides. Biosci. Biotechnol. Biochem. **58**: 554–555.

Mohd Azhar, S.H., Abdulla, R., Jambo, S.A., Marbawi, H., Gansau, J.A., Mohd Faik, A.A., and Rodrigues, K.F. (2017). Yeasts in sustainable bioethanol production: A review. Biochem. Biophys. Reports **10**: 52–61.

Moses, T. et al. (2015a). OSC2 and CYP716A14v2 Catalyze the Biosynthesis of Triterpenoids for the Cuticle of Aerial Organs of *Artemisia annua*. Plant Cell Online **27**: 286–301.

Moses, T., Papadopoulou, K.K., and Osbourn, A. (2014a). Metabolic and functional diversity of saponins, biosynthetic intermediates and semi-synthetic derivatives. Crit. Rev. Biochem. Mol. Biol. **49**: 439–462.

Moses, T., Pollier, J., Almagro, L., Buyst, D., Van Montagu, M., Pedreño, M. a, Martins, J.C., Thevelein, J.M., and Goossens, A. (2014b). Combinatorial biosynthesis of sapogenins and saponins in Saccharomyces cerevisiae using a C-16α hydroxylase from Bupleurum falcatum. Proc. Natl. Acad. Sci. U. S. A. **111**: 1634–9.

Moses, T., Pollier, J., Faizal, A., Apers, S., Pieters, L., Thevelein, J.M., Geelen, D., and Goossens, A. (2015b). Unraveling the triterpenoid saponin biosynthesis of the african shrub maesa lanceolata. Mol. Plant **8**: 122–135.

Moses, T., Pollier, J., Thevelain, J.M., and Goossens, A. (2013). Bioengineering of plant (tri) terpenoids: from metabolic engineering of plants to synthetic biology in vivo and in vitro. New Phytol. **200**: 27–43.

Moses, T., Thevelein, J.M., Goossens, A., and Pollier, J. (2014c). Comparative analysis of CYP93E proteins for improved microbial synthesis of plant triterpenoids. Phytochemistry **108**: 47–56.

Mugford, S.T. et al. (2009). A Serine Carboxypeptidase-Like Acyltransferase Is Required for Synthesis of Antimicrobial Compounds and Disease Resistance in Oats. Plant Cell Online **21**: 2473–2484.

Nag, S.A. (2012). Ginsenosides as anticancer agents: In vitro and in vivo activities, structure–activity relationships, and molecular mechanisms of action. Front. Pharmacol. **3**: 25.

Naoumkina, M. a, Modolo, L. V, Huhman, D. V, Urbanczyk-Wochniak, E., Tang, Y., Sumner, L.W., and Dixon, R. a (2010). Genomic and coexpression analyses predict multiple genes involved in triterpene saponin biosynthesis in Medicago truncatula. Plant Cell **22**: 850–866.

Nelson, D. and Werck-Reichhart, D. (2011). A P450-centric view of plant evolution. Plant J. **66**: 194–211.

Nelson, D.R. (2009). The Cytochrome P450 Homepage. Hum. Genomics **4**: 59–65.

Nicolaou, K.C., Sorensen, E.J., and Winssinger, N. (1998). The Art and Science of Organic and Natural Products Synthesis. J. Chem. Educ. **75**: 1225.

Nielsen, J.K., Nagao, T., Okabe, H., and Shinoda, T. (2010). Resistance in the plant, Barbarea vulgaris, and counter-adaptations in flea beetles mediated by saponins. J. Chem. Ecol. **36**: 277–285.

Nisar, N., Li, L., Lu, S., Khin, N.C., and Pogson, B.J. (2015). Carotenoid metabolism in plants. Mol. Plant **8**: 68–82.

Oka, T. and Jigami, Y. (2006). Reconstruction of de novo pathway for synthesis of UDP-glucuronic acid and UDP-xylose from intrinsic UDP-glucose in Saccharomyces cerevisiae. FEBS J. **273**: 2645–2657.

Osbourn, A., Goss, R.J., and Field, R.A. (2011). The saponins: polar isoprenoids with important and diverse biological activities. Nat Prod Rep **28**: 1261–1268.

Osbourn, A.E. and Field, B. (2009). Operons. Cell. Mol. Life Sci. **66**: 3755–3775.

Paddon, C.J. et al. (2013). High-level semi-synthetic production of the potent antimalarial artemisinin. Nature **496**: 528–532.

Paddon, C.J. and Keasling, J.D. (2014). Semi-synthetic artemisinin: a model for the use of synthetic biology in pharmaceutical development. Nat Rev Micro **12**: 355–367.

Papadopoulou, K., Melton, R.E., Leggett, M., Daniels, M.J., and Osbourn, a E. (1999). Compromised disease resistance in saponin-deficient plants. Proc. Natl. Acad. Sci. U. S. A. **96**: 12923–12928.

Peterson, J.A., Ebel, R.E., O'Keeffe, D.H., Matsubara, T., and Estabrook, R.W. (1976). Temperature Dependence of Cytochrome P-450 Reduction. A Model for NADPH-Cytochrome P-450 Reductase:Cytochrome P-450 Interaction. J. Biol. Chem. **251**: 4010–4016.

Polakowski, T., Stahl, U., and Lang, C. (1998). Overexpression of a cytosolic hydroxymethylglutaryl-CoA reductase leads to squalene accumulation in yeast.

Appl. Microbiol. Biotechnol. **49**: 66–71.

Pollier, J. and Goossens, A. (2012). Oleanolic acid. Phytochemistry **77**: 10–15.

Poralla, K. (1994). The possible role of a repetitive amino acid motif in evolution of triterpenoid cyclases. Bioorganic Med. Chem. Lett. **4**: 285–290.

Qi, X., Bakht, S., Leggett, M., Maxwell, C., Melton, R., and Osbourn, A. (2004). A gene cluster for secondary metabolism in oat: implications for the evolution of metabolic diversity in plants. Proc. Natl. Acad. Sci. U. S. A. **101**: 8233–8.

Ragupathi, G., Gardner, J.R., Livingston, P.O., and Gin, D.Y. (2011). Natural and synthetic saponin adjuvant QS-21 for vaccines against cancer. Expert Rev. Vaccines **10**: 463–470.

Rajkumar, A.S., Liu, G., Bergenholm, D., Arsovska, D., Kristensen, M., Nielsen, J., Jensen, M.K., and Keasling, J.D. (2016). Engineering of synthetic, stress-responsive yeast promoters. Nucleic Acids Res. **44**: 1–12.

Redden, H. and Alper, H.S. (2015). The development and characterization of synthetic minimal yeast promoters. Nat. Commun. **6**: 1–9.

Reed, J.R. and Backes, W.L. (2012). Formation of P450 • P450 complexes and their effect on P450 function. Pharmacol. Ther. **133**: 299–310.

Renault, H., Bassard, J.E., Hamberger, B., and Werck-Reichhart, D. (2014). Cytochrome P450-mediated metabolic engineering: Current progress and future challenges. Curr. Opin. Plant Biol. **19**: 27–34.

Ro, D.K. et al. (2006). Production of the antimalarial drug precursor artemisinic acid in engineered yeast. Nature **440**: 940–943.

Ryan, O.W., Skerker, J.M., Maurer, M.J., Li, X., Tsai, J.C., Poddar, S., Lee, M.E., DeLoache, W., Dueber, J.E., Arkin, A.P., and Cate, J.H.D. (2014). Selection of chromosomal DNA libraries using a multiplex CRISPR system. Elife **3**: 1–15.

Saimaru, H., Orihara, Y., Tansakul, P., Kang, Y.-H., Shibuya, M., and Ebizuk, Y. (2007). Production of Triterpene Acids by Cell-suspension Cultures of Olea europaea. Chem. Pharm. Bull. **55**: 784–788.

Salmon, M., Thimmappa, R.B., Minto, R.E., Melton, R.E., Hughes, R.K., O'Maille, P.E., Hemmings, A.M., and Osbourn, A. (2016). A conserved amino acid residue critical for product and substrate specificity in plant triterpene synthases. Proc. Natl. Acad. Sci. **113**: E4407–E4414.

Santos-Rosa, H., Leung, J., Grimsey, N., Peak-Chew, S., and Siniossoglou, S. (2005). The yeast lipin Smp2 couples phospholipid biosynthesis to nuclear membrane growth. EMBO J. **24**: 1931–1941.

Sayama, T. et al. (2012). The Sg-1 Glycosyltransferase Locus Regulates Structural Diversity of Triterpenoid Saponins of Soybean. Plant Cell **24**: 2123–2138.

Schückel, J., Rylott, E.L., Grogan, G., and Bruce, N.C. (2012). A Gene-Fusion Approach to Enabling Plant Cytochromes P450 for Biocatalysis. ChemBioChem **13**: 2758–2763.

Seki, H., Ohyama, K., Sawai, S., Mizutani, M., Ohnishi, T., Sudo, H., Akashi, T., Aoki, T., Saito, K., and Muranaka, T. (2008). Licorice beta-amyrin 11-oxidase, a cytochrome P450 with a key role in the biosynthesis of the triterpene sweetener glycyrrhizin. Proc. Natl. Acad. Sci. U. S. A. **105**: 14204–14209.

Seki, H., Sawai, S., Ohyama, K., Mizutani, M., Ohnishi, T., Sudo, H., Fukushima, E.O., Akashi, T., Aoki, T., Saito, K., and Muranaka, T. (2011). Triterpene functional genomics in licorice for identification of CYP72A154 involved in the biosynthesis of glycyrrhizin. Plant Cell **23**: 4112–23.

Sezutsu, H., le Goff, G., and Feyereisen, R. (2013). Origins of P450 diversity. Philos. Trans. R. Soc. B Biol. Sci. **368**.

Shakhidoyatov, K.M., Rashkes, A.M., and Khidyrova, N.K. (1997). Components of Cottonplant Leaves, Their Functional Role and Biological Activity. Chem. Nat. Compd. **33**: 605–616.

Shan, H., Segura, M.J.R., Wilson, W.K., Lodeiro, S., and Matsuda, S.P.T. (2005). Enzymatic cyclization of dioxidosqualene to heterocyclic triterpenes. J. Am. Chem. Soc. **127**: 18008–18009.

Shibuya, M., Nishimura, K., Yasuyama, N., and Ebizuka, Y. (2010). Identification and characterization of glycosyltransferases involved in the biosynthesis of soyasaponin I in Glycine max. FEBS Lett. **584**: 2258–2264.

Shinoda, T., Nagao, T., Nakayama, M., Serizawa, H., Koshioka, M., Okabe, H., and Kawai, A. (2002). Identification of a triterpenoid saponin from a crucifer, Barbarea vulgaris, as a feeding deterrent to the diamondback moth, Plutella xylostella. J. Chem. Ecol. **28**: 587–599.

Sievers, F. and Higgins, D.G. (2018). Clustal Omega for making accurate alignments of many protein sequences. Protein Sci. **27**: 135–145.

Smith, P.F., Ogundele, A., Forrest, A., Wilton, J., Salzwedel, K., Doto, J., Allaway, G.P., and Martin, D.E. (2007). Phase I and II study of the safety, virologic effect, and pharmacokinetics/pharmacodynamics of single-dose 3-O-(3′3′- dimethylsuccinyl)betulinic acid (bevirimat) against human immunodeficiency virus Infection. Antimicrob. Agents Chemother. **51**: 3574–3581.

Strawser, C., Schadt, K., Hauser, L., McCormick, A., Wells, M., Larkindale, J., Lin, H., and Lynch, D.R. (2017). Pharmacological therapeutics in Friedreich ataxia: the present state. Expert Rev. Neurother. **17**: 895–907.

Suh, N., Paul, S., Lee, H.J., Yoon, T., Shah, N., Son, A.I., Reddi, A.H., Medici, D., and Sporn, M.B. (2012). Synthetic triterpenoids, CDDO-Imidazolide and CDDO-Ethyl amide, induce chondrogenesis. Osteoarthr. Cartil. **20**: 446–450.

Sun, H.X., Xie, Y., and Ye, Y.P. (2009). Advances in saponin-based adjuvants. Vaccine **27**: 1787–1796.

Suzuki, H., Achnine, L., Xu, R., Matsuda, S.P.T., and Dixon, R. a (2002). A genomics approach to the early stages of triterpene saponin biosynthesis in

Medicago truncatula. Plant J **32**: 1033–1048.

Tamura, K., Teranishi, Y., Ueda, S., Suzuki, H., Kawano, N., Yoshimatsu, K., Saito, K., Kawahara, N., Muranaka, T., and Seki, H. (2017). Cytochrome P450 monooxygenase CYP716A141 is a unique β-amyrin C-16β oxidase involved in triterpenoid saponin biosynthesis in Platycodon grandiflorus. Plant Cell Physiol. **58**: 874–884.

Teste, M.A., Duquenne, M., François, J.M., and Parrou, J.L. (2009). Validation of reference genes for quantitative expression analysis by real-time RT-PCR in Saccharomyces cerevisiae. BMC Mol. Biol. **10**: 99.

Thimmappa, R., Geisler, K., Louveau, T., O'Maille, P., and Osbourn, A. (2014). Triterpene Biosynthesis in Plants. Annu. Rev. Plant Biol. **65**: 225–257.

Thoma, R., Schulz-Gasen, T., D'Arcy, B., Benz, J., Aebi, J., Dehmlow, H., Hennig, M., Stihle, M., and Ruf, A. (2004). Insight into steroid scaffold formation from the structure of human oxidosqualene cyclase. Nature **432**: 118–122.

Timson, D.J. (2007). Galactose Metabolism in Saccharomyces cerevisiae. Dyn. Biochem. Process Biotechnol. Mol. Biol.: 63–73.

Tong, X., Han, L., Duan, H., Cui, Y., Feng, Y., Zhu, Y., Chen, Z., and Yang, S. (2017). The derivatives of Pulsatilla saponin A, a bioactive compound from Pulsatilla chinensis: Their synthesis, cytotoxicity, haemolytic toxicity and mechanism of action. Eur. J. Med. Chem. **129**: 325–336.

Tsao, T. et al. (2010). Role of peroxisome proliferator-activated receptor-gamma and its coactivator DRIP205 in cellular responses to CDDO (RTA-401) in acute myelogenous leukemia. Cancer Res. **70**: 4949–60.

Urban, P., Mignotte, C., Kazmaier, M., Delorme, F., and Pompon, D. (1997). Cloning, Yeast Expression, and Characterization of the Coupling of Two Distantly Related Arabidopsis thaliana NADPH-Cytochrome P450 Reductases with P450 CYP73A5. Biochemistry **272**: 19176–19186.

Urban, P., Werck-Reichhart, D., Teutsch, H.G., Durst, F., Regnier, S., Kazmaier, M., and Pompon, D. (1994). Characterization of recombinant plant cinnamate 4-hydroxylase produced in yeast: Kinetic and spectral properties of the major plant P450 of the phenylpropanoid pathway. Eur. J. Biochem. **222**: 843–850.

Veen, M., Stahl, U., and Lang, C. (2003). Combined overexpression of genes of the ergosterol biosynthetic pathway leads to accumulation of sterols in. FEMS Yeast Res. **4**: 87–95.

Vincken, J.P., Heng, L., de Groot, A., and Gruppen, H. (2007). Saponins, classification and occurrence in the plant kingdom. Phytochemistry **68**: 275–297.

Wang, P., Wei, Y., Fan, Y., Liu, Q., Wei, W., Yang, C., Zhang, L., Zhao, G., Yue, J., Yan, X., and Zhou, Z. (2015). Production of bioactive ginsenosides Rh2 and Rg3 by metabolically engineered yeasts. Metab. Eng. **29**: 97–105.

Wang, Z., Guhling, O., Yao, R., Li, F., Yeats, T.H., Rose, J.K.C., and Jetter, R. (2011). Two Oxidosqualene Cyclases Responsible for Biosynthesis of Tomato Fruit Cuticular Triterpenoids. Plant Physiol. **155**: 540–552.

Wei, W., Wang, P., Wei, Y., Liu, Q., Yang, C., Zhao, G., Yue, J., Yan, X., and Zhou, Z. (2015). Characterization of Panax ginseng UDP-glycosyltransferases catalyzing protopanaxatriol and biosyntheses of bioactive ginsenosides F1 and Rh1 in metabolically engineered yeasts. Mol. Plant **8**: 1412–1424.

Wei, X., Zhang, X., Wu, Q., Wang, H., Shen, D., Qiu, Y., Song, J., and Li, X. (2012). Cloning, Characterization and Real-time RT-PCR Analysis of a Key Gene beta-Amyrin synthase for Saponin Biosynthesis in Barbarea vulgaris. Acta Hortic. Sin. **39**: 923–930.

Weinhandl, K., Winkler, M., Glieder, A., and Camattari, A. (2014). Carbon source dependent promoters in yeasts. Microb. Cell Fact. **13**: 1–17.

Wendt, K.U., Poralla, K., and Schulz, G.E. (1997). Structure and function of a squalene cyclase. Science (80-.). **277**: 1811–1815.

Werck-Reichhart, D. and Feyereisen, R. (2000). Cytochromes P450: a success story. Genome Biol. **1**: REVIEWS3003.

Williams, L.A.D. (1999). *Rhizophora mangle* (Rhizophoraceae) triterpenoids with insecticidal activity. Naturwissenschaften **86**: 450–452.

Xiao, G. and Yu, B. (2013). Total synthesis of starfish saponin goniopectenoside B. Chem. - A Eur. J. **19**: 7708–7712.

Xu, G., Cai, W., Gao, W., and Liu, C. (2016). A novel glucuronosyltransferase has an unprecedented ability to catalyse continuous two-step glucuronosylation of glycyrrhetinic acid to yield glycyrrhizin. New Phytol. **212**: 123–135.

Yan, X., Fan, Y., Wei, W., Wang, P., Liu, Q., Wei, Y., Zhang, L., Zhao, G., Yue, J., and Zhou, Z. (2014). Production of bioactive ginsenoside compound K in metabolically engineered yeast. Cell Res. **24**: 770–773.

Yang, X.D., Yang, Y.Y., Ouyang, D.S., and Yang, G.P. (2015). A review of biotransformation and pharmacology of ginsenoside compound K. Fitoterapia **100**: 208–220.

Yasumoto, S., Fukushima, E.O., Seki, H., and Muranaka, T. (2016). Novel triterpene oxidizing activity of Arabidopsis thaliana CYP716A subfamily enzymes. **590**: 533–540.

Yasumoto, S., Seki, H., Shimizu, Y., Fukushima, E.O., and Muranaka, T. (2017). Functional Characterization of CYP716 Family P450 Enzymes in Triterpenoid Biosynthesis in Tomato. Front. Plant Sci. **8**.

Zainal, B., Abdah, M.A., Taufiq-Yap, Y.H., Roslida, A.H., and Rosmin, K. (2014). Anticancer Agents from Non-Edible Parts of Theobroma cacao. Nat. Prod. Chem. Res. **2**.

Zangar, R.C., Davydov, D.R., and Verma, S. (2004). Mechanisms that regulate

production of reactive oxygen species by cytochrome P450. Toxicol. Appl. Pharmacol. **199**: 316–331.

Zhang, G., Cao, Q., Liu, J., Li, J., and Li, C. (2015). Refactoring β-Amyrin Synthesis in S. cerevisiae. AIChE J. **61**: 3172–3179.

Zhao, F., Bai, P., Liu, T., Li, D., Zhang, X., Lu, W., and Yuan, Y. (2016). Optimization of a cytochrome P450 oxidation system for enhancing protopanaxadiol production in Saccharomyces cerevisiae. Biotechnol. Bioeng. **113**: 1787–1795.

Zhao, F., Du, Y., Bai, P., Liu, J., Lu, W., and Yuan, Y. (2017). Enhancing Saccharomyces cerevisiae reactive oxygen species and ethanol stress tolerance for high-level production of protopanoxadiol. Bioresour. Technol. **227**: 308–316.

Zhao, Y., Fan, J., Wang, C., Feng, X., and Li, C. (2018). Enhancing oleanolic acid production in engineered Saccharomyces cerevisiae. Bioresour. Technol. **257**: 339–343.

Zhou, C., Li, J., Li, C., and Zhang, Y. (2016). Improvement of betulinic acid biosynthesis in yeast employing multiple strategies. BMC Biotechnol. **16**: 1–9.

Zhu, C., Tang, P., and Yu, B. (2008). Total synthesis of potent antitumor agent (-)-lasonolide A: A cycloaddition-based strategy. J. Am. Acad. Dermatol. **130**: 5872–5873.

Zhu, M., Wang, C., Sun, W., Zhou, A., Wang, Y., Zhang, G., Zhou, X., Huo, Y., and Li, C. (2018). Boosting 11-oxo-β-amyrin and glycyrrhetinic acid synthesis in Saccharomyces cerevisiae via pairing novel oxidation and reduction system from legume plants. Metab. Eng. **45**: 43–50.

Zhuang, Y., Yang, G.Y., Chen, X., Liu, Q., Zhang, X., Deng, Z., and Feng, Y. (2017a). Biosynthesis of plant-derived ginsenoside Rh2 in yeast via repurposing a key promiscuous microbial enzyme. Metab. Eng. **42**: 25–32.

Zhuang, Y., Yang, G.Y., Chen, X., Liu, Q., Zhang, X., Deng, Z., and Feng, Y. (2017b). Biosynthesis of plant-derived ginsenoside Rh2 in yeast via repurposing a key promiscuous microbial enzyme. Metab. Eng. **42**: 25–32.